DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

ECONOMIC VALUE OF EWA LITE: A FUNCTIONAL THEORY OF LEARNING IN GAMES

Teck-Hua Ho
The Wharton School, University of Pennsylvania

Colin F. Camerer
California Institute of Technology

Juin-Kuan Chong
National University of Singapore

SOCIAL SCIENCE WORKING PAPER 1122

May 2001

Economic Value of EWA Lite: A Functional Theory of Learning in Games

Teck-Hua Ho Colin F. Camerer Juin-Kuan Chong

Abstract

EWA Lite is a one-parameter theory of learning in normal-form games. It approximates the free parameters in an earlier model (EWA) with functions of experience. The theory is tested on seven different games and compared to other learning and equilibrium theories. Either EWA Lite or parameterized EWA predict best, but one kind of reinforcement learning predicts well in games with mixed-strategy equilibrium. Belief learning models fit worst. The economic value of theories is measured by how much more subjects would have earned if they followed theory recommendations. EWA Lite and EWA add the most economic value in every game but one.

*This paper should not be circulated or quoted without permission. Thanks to participants in the Southern Economics Association meetings (December, 2000), the Wharton School Decision Processes Workshop, and C. Monica Capra for comments.

"In nature hybrid species are usually sterile, but in science the reverse is true"– Francis Crick (1988, p. 150)

The power of equilibrium models comes from their ability to produce precise predictions using only the structure of a game and some assumptions about rationality. Statistical models of learning should strive to be almost as parsimonious and precise, while also predicting the time path of actual observations more accurately than equilibrium theories do. Most learning models do this by specifying a formula for predicting future choices from past experiences (often at the population level), using one or more free parameters which are typically estimated from data. This paper describes a theory of learning in decisions and games called EWA Lite, with only one parameter. EWA Lite predicts the time path of individual behavior in any normal-form game (given initial conditions) including new games in which behavior has never been observed.

The key innovation in EWA Lite is the replacement of parameter values with functions of players' experience, which can vary across games, individuals, and time periods. Replacing parameters with functions kills two birds with one stone. The first bird is explaining why estimated model parameter values vary significantly across games (as earlier research showed). The functions in EWA Lite reproduce these cross-game differences endogeneously, as a function of the interaction between experience and game structure. The second "bird" is parsimony. By replacing parameters with functions, in EWA Lite only one free parameter needs to be estimated or fixed a priori. (The parameter captures sensitivity of players to differences in numerical ratings of strategies; it is probably impossible to fit data, and hence have a zero-parameter theory, without it.[1]) There are other one-parameter theories but they do not predict as well across games as EWA Lite does.

EWA Lite is used to fit and predict data from seven experimental data sets, and is compared to general versions of belief and reinforcement learning, and quantal response equilibrium. In out-of-sample forecasting, either EWA Lite or its parameterized precursor, EWA, tend to predict best, although a new version of reinforcement predicts as well in some cross-game forecasting. We also introduce a new criterion for judging usefulness

[1]If the goal is to predict the most likely choice, EWA Lite can be reduced to a zero-parameter theory by setting the experience weight $N(0)$ to 0 (see below).

of theories– economic value. The economic value of a theory is measured by how well model forecasts of behavior of other players would improve a player's profitability of best responses to those forecasts were substituted for the player's actual choices. EWA and EWA Lite have the highest economic value in 5 to 7 (depending on what parameters were used to generate forecasts) of the 7 games we study.

While EWA Lite is rather narrowly focused to explain learning in normal-form games, sensible extensions of it can be applied to field settings such as evolution of economic institutions (e.g., internet auctions or pricing), investors and policymakers learning about equity market fluctuations or macroeconomic phenomena, and consumer choice. A variant of the EWA theory is used by Teck-Hua Ho and Juin-Kuan Chong (1999) to fit and predict 130,000 product choices by consumers.[2] Readers who are interested in learning in field settings should be interested in how subjects learn in experimental games because understanding learning in the lab will surely help us understand learning in the field.

1 EWA learning and its limitations

In earlier work we proposed a model of learning called experience-weighted attraction (EWA) theory (Colin Camerer and Ho 1998, 1999). Learning in EWA is characterized by changes in (unobserved) attractions based on experience. EWA was designed to be a gene-splice or hybrid of two models, reinforcement and belief learning, which have been used to study learning in games. The EWA model wraps a parametric skin around both of those theories, which are historically-interesting special cases on the boundary of the parameter space.

Attractions determine the probabilities of choosing different strategies through a logistic response function. For player i, there are m_i strategies (indexed by j) which have initial attractions denoted $A_i^j(0)$ (either estimated as free parameters from the data, specified by some theory of initial conditions, or "burned in" using the first period data). Denote i's j'th strategy by s_i^j, chosen strategies by i and other players (denoted $-i$)

[2]Their theory uses 80% *fewer* parameters than the leading theory used in marketing and predicts 20% better.

by $s_i(t)$ and $s_{-i}(t)$, and player i's payoffs by $\pi_i(s_i^j, s_{-i}(t))$. Define an indicator function $I(x, y)$ to be zero if $x \neq y$ and one if $x = y$. The EWA attraction updating equation is[3]

$$A_i^j(t) = \frac{\phi \cdot N(t-1) \cdot A_i^j(t-1) + [\delta + (1-\delta) \cdot I(s_i^j, s_i(t))] \cdot \pi_i(s_i^j, s_{-i}(t))}{N(t-1) \cdot \phi \cdot (1-\kappa) + 1} \tag{1.1}$$

and the experience weight is updated according to $N(t) = N(t-1) \cdot \phi(1-\kappa) + 1$.

The parameter δ is the weight placed on foregone payoffs. It presumably is affected by imagination (in psychological terms, the strength of counterfactual simulation) and reliability of information about foregone payoffs (Dana Heller and Rajiv Sarin, 2000). The parameter ϕ reflects decay of previous attractions due to forgetting or to deliberate ignorance of old experience when the learning environment is changing. The parameter κ controls the rate at which attractions grow. When $\kappa = 0$ attractions are weighted averages of reinforcements and decayed lagged attractions; when $\kappa = 1$ attractions cumulate. The growth rate of attractions is important because in the logit model the difference in attractions determines the spread of choice probabilities. The initial experience weight $N(0)$ is like a strength of prior beliefs and is estimated using data. Since it usually plays a minor role in predicting learning, we restrict $N(0) = 1$ in our specification of EWA Lite.[4]

[3]This updating equation assumes that subjects know the payoffs of strategies that were not chosen. In Camerer, Ho, and Xin Wang (1999), we apply EWA model to games where such payoffs are not available by allowing subjects to learn about them through experience. See Yan Chen and Yuri Khoroshilov (2000) for a similar extension.

[4]In earlier work we imposed the restriction $N(0) < \frac{1}{\phi \cdot (1-\kappa)}$ so that the experience weight, which is updated according to $N(t) = \phi \cdot (1-\kappa) \cdot N(t-1) + 1$, is always increasing. When the inequality binds, as it often does, $N(0)$ and κ are not separately identified. We also switched notation in the denominator (previously we denoted the product $\phi \cdot (1-\kappa)$ by a single variable ρ). The notation shift makes the way in which κ controls cumulation versus averaging more transparent.

A logit response function[5] is used to map attractions into probabilities:

$$P_i^j(t+1) = \frac{e^{\lambda \cdot A_i^j(t)}}{\sum_{k=1}^{m_i} e^{\lambda \cdot A_i^k(t)}} \tag{1.2}$$

where λ is the response sensitivity.

A key insight from our earlier work is that reinforcement and belief learning approaches are closely related in an interesting way.[6] When $\delta = 0$ the EWA rule is the same as models in which only chosen strategies are reinforced, originating in studies of animal learning. When $\kappa = 1$ the rule is a simpler form of cumulative reinforcement model studied by Calvin Harley (1981) and Alvin Roth and Ido Erev (1995) (see also Robert Bush and Frederick Mosteller, 1955; John Cross, 1983; Patrick McAllister, 1991; Brian Arthur, 1991). When $\kappa = 0$ the rule is like the averaging reinforcement model of Roth, Greg Barron, Erev and Robert Slonim (2000).

When $\delta = 1$ and $\kappa = 0$ the EWA rule is equivalent to belief learning using weighted fictitious play (Fudenberg and Levine, 1998). The EWA rule shows that belief learning is *not* fundamentally different than reinforcement. It is a kind of reinforcement in which unchosen strategies are reinforced as strongly as chosen ones, and attractions are averages.

One way to see the relation of different learning rules is a cube showing configurations of parameter values (see Figure 1). Each point in the cube is a triple of parameter values which specifies a precise updating equation (leaving aside λ and initial conditions). The cube shows the EWA family of learning rules. Corners and vertices of the EWA cube correspond to boundary special cases. The corner of the cube with $\phi = \kappa = 0, \delta = 1$, is Cournot best-response dynamics. The corner $\kappa = 0, \phi = \delta = 1$, is standard fictitious play (George Brown, 1951 and Julia Robinson, 1951). The edge connecting these corners, $\delta = 1, \kappa = 0$, is the class of weighted fictitious play rules (e.g., Fudenberg and Levine,

[5] In previous work (Camerer and Ho, 1998) we compared the logit rule and a power function in which attractions are exponentiated and normalized $P_i^j(t+1) = (A_i^j(t))^\lambda / \sum_{k=1}^{m_i}(A_i^k(t))^\lambda$. The logit rule fits slightly better, depending on how fit is evaluated. The denominator of the attraction updating equation $(\phi \cdot (1-\kappa)+1)$ divides out when using the power rule, which conserves a degree of freedom and may be useful for some purposes.

[6] See also Yin-Wong Cheung and Daniel Friedman, 1997, pp. 54-55; Drew Fudenberg and David Levine, 1998, pp. 1084-1085; Ed Hopkins, 1999.

1998). The edges with $\delta = 0$ and κ equal to zero or one are averaging and cumulative choice reinforcement rules.

The EWA cube is a visual aid to show the relations and differences among theories. But EWA is also a bet that psychologically plausible learning rules have parameter values which are in the interior of the cube rather than on vertices and corners. (That is, as Francis Crick suggested in the quote that opened this paper, a scientific hybrid may work better.) Reinforcement theories with $\delta = 0$ ignore foregone payoffs entirely.[7] Belief learning using weighted fictitious play ($\delta = 1$) ignores the difference between received and foregone payoffs, which is also unlikely.[8] Put differently, reinforcement models assume that received payoffs matter more than foregone payoffs ($\delta < 1$) and belief learning says that foregone payoffs do matter ($\delta > 0$). Both intuitions are plausible and EWA includes both if δ is between zero and one. Intermediate estimated values of δ could result if some subjects learn according to reinforcement and others according to weighted fictitious play, but direct tests allowing "latent class" heterogeneity show this is not so (Camerer and Ho, 1998).

Estimates by ourselves and others (see Camerer, David Hsia, and Ho, 2000) have shown in 31 data sets that EWA generally predicts (out of sample) more accurately than the special cases of reinforcement and weighted fictitious play, except in games with mixed-strategy equilibrium (where all models only improve a little on Nash equilibrium). However, EWA has been criticized for having too many free parameters. EWA Lite answers this criticism because it has only one parameter, so it is simpler than most reinforcement and belief models.

Figure 1 also shows estimated parameter triples from twenty games (see Camerer, Hsia, and Ho, 2000). Each point corresponds to a different game. If one of the special case theories is a good approximation to how people generally behave across games, the parameters will cluster in the corner or vertex corresponding to that theory. In fact,

[7]This assumption is implausible when foregone payoffs are known, and has been rejected by studies comparing different information conditions, e.g., Dilip Mookerjhee and Barry Sopher, 1994, and Amnon Rapoport and Erev, 1998; and John Van Huyck, Raymond Battalio and Frederick Rankin, 1996.

[8]There is substantial evidence that people underweight opportunity costs compared to out-of-pocket costs (e.g., Daniel Kahneman, Jack Knetsch, and Richard Thaler, 1991). Since the difference between foregone and received payoffs is an opportunity cost (or gain) if it is underweighted then $\delta < 1$.

parameters tend to be sprinkled around the cube. Estimates from coordination games usually have high values of δ and κ. Estimates from mixed games tend to have low δ and κ (close to the averaging reinforcement corner with ϕ close to one).

A second criticism of EWA is that estimated parameter values vary across games (although parametric variation is characteristic of all learning models, e.g., Vincent Crawford, 1995; Cheung and Friedman, 1997). Roth, et al. (2000) note that our earlier work found "very different parameters in, apparently, very similar constant sum games. Their [Camerer and Ho's] research leads to the pessimistic conclusion that, at least currently, it is impossible to predict behavior in a new situation". Their conclusion *is* too pessimistic, because EWA Lite maps games to parameters and makes it possible to predict new games.

2 EWA Lite

EWA Lite replaces the three central parameters of EWA, ϕ, δ, κ with deterministic functions $\phi_i(t), \delta_i(t), \kappa_i(t)$ of player i's experience up to period t. These functions determine parameter values for each player and period, which are then plugged into the EWA updating equation to determine attractions. Updated attractions determine choice probabilities according to the logit rule, given a value of λ. Standard methods for optimizing fit given λ can then be used to find which λ fits best.

2.1 The change-detector function $\phi_i(t)$

The decay rate ϕ is sometimes interpreted as forgetting, an interpretation carried over from reinforcement models of animal learning. Certainly forgetting does occur, but the more important variation in $\phi_i(t)$ across games is probably a player's perception of how quickly the learning environment is changing. The function $\phi_i(t)$ should therefore "detect change". As in physical change detectors (e.g., security systems or smoke alarms), the challenge is to detect change when it is really occurring, but not falsely mistake noise for change too often. The core of the function is a "surprise index", the difference between the other players' strategies in the window of the last W periods and the average strategy of others in all previous periods (where W is the minimal support of Nash equilibria). We specify the function in terms of relative frequencies of strategies, without using information about how strategies are ordered, so it can be applied to non-ordered strategies (e.g., rows in a normal-form game). The change-detection function $\phi_i(t)$ is

$$\phi_i(t) = 1 - .5(\sum_{j=1}^{m_{-i}}[\frac{\sum_{\tau=t-W+1}^{t} I(s_{-i}^j, s_{-i}(\tau))}{W} - \frac{\sum_{\tau=1}^{t} I(s_{-i}^j, s_{-i}(\tau))}{t}]^2) \tag{2.1}$$

The term $\frac{\sum_{\tau=t-W+1}^{t} I(s_{-i}^j, s_{-i}(\tau))}{W}$ is the j-th element of a vector that simply counts how often strategy j was played by the others in periods $t-W+1$ to t, and divides by W. The term $\frac{\sum_{\tau=1}^{t} I(s_{-i}^j, s_{-i}(\tau))}{t}$ is the relative frequency count of the j-th strategy over all t periods. [9] To measure change, we take the differences in corresponding elements of the two frequency vectors, square them, and sum over strategies. Since the maximum difference is two, the function is normalized by dividing the sum of squared differences by two, and subtracting the normalized figure from one. When recent observations of what others have done deviate a lot from all previous observations, the deviations in strategy frequencies will be high and ϕ will be low. When recent observations are like old observations, ϕ will be high.

The key to modeling ϕ is keeping it close to one unless there is an unmistakably persistent change in what others are doing. It is dangerous to let ϕ become too low because doing so erases everything that has been learned, by giving a low weight to previous attractions which summarize previous experience. The $\phi_i(t)$ function dips lowest in the extreme case in which one strategy is played until $t-1$, and then a new strategy is played. Then $\phi_i(t)$ is $\frac{2t-1}{t^2}$. This expression declines gracefully toward zero as the string of identical choices up to period t grows longer. (For t=2, 3, and 10 the $\phi_i(t)$ values are .75, .56, and .19.) This makes sense because after more identical choices in a row a new choice is a bigger surprise. A lower ϕ puts a higher decay on old experience. Another interesting special case is when different strategies have been played in every period up to $t-1$, and another different strategy is played in period t. (This is often true in games

[9] In the case of games with multiple players, frequency count of the relevant aggregate statistics is used. For example, in median action game, frequency count of the median strategy by all other players in each period is used.

with large strategy spaces, such as p-beauty contests, when order of strategies is not used.) Then $\phi_i(t) = .5 + \frac{1}{2t}$, which starts at .75 and asymptotes at .5 as t increases.

So far we have neglected an important detail: What's W? W is the smallest support of all the Nash equilibria (the number of strategies played with positive probability). In games with a pure strategy equilibrium, $W = 1$. In games with mixed equilibria W is larger than one. In these games, a certain amount of period to period change is expected. The number of strategies with positive probability, W, tells us roughly "how much" variation to expect, and hence, how many previous observations to average to smooth perceived change.

2.2 Responsiveness $\delta_i(t)$

The parameter δ is the weight on foregone payoffs. We tried several specifications[10] and settled on the following: $\delta_i(t) = \phi_i(t)/W$.

The specification $\delta = \frac{\phi}{W}$ is appealing for three reasons. First, tying δ to the change measure ϕ recognizes the fact that best-responding to foregone payoffs is a good strategy when the environment is stable, so that δ should be near one when ϕ is near one. But

[10]We tried obvious routes like setting $\delta = 1$ (as in belief learning models). This specification does poorly in mixed games in which estimated δs are close to zero. We also explored specifications which used the person's own history of responses to judge whether she is responsive to foregone payoffs or not and customize different δ values for each player. Suppose a player moves toward a worse response after period t and anticipate earning less from their strategy in $t+1$, given what the other player did last time, than if they had simply repeated their previous choice (i.e., $\pi_i(s_i(t+1), s_{-i}(t)) < \pi_i(s_i(t), s_{-i}(t)))$). When this happens we give the player a score of 0 in period $t+1$ (i.e., they were not responsive to foregone payoffs in that period). If the player moves toward a (weakly) better response, or repeats her choice when it is the ex post best response, she gets a responsiveness score of 1. The player's δ is the average of these 0 and 1 scores over all previous periods. This approach uses the person's own behavior to identify their "type", a la "revealed preference" in consumer theory. This approach does capture an important cross-game regularity, which is that players are more responsive in games with pure equilibria than in games with mixed equilibria so δ's are lower in mixed games. (When there are mixed equilibria players often do not best-respond to the last choice by their opponent, which leads to low values of δ.) However, this responsiveness-indexing approach just did not fit that well. Without stronger intuitions, axiomatic underpinnings, or earlier research to guide us, we opted for the simpler $\delta = \phi/W$ specification. Researchers looking for ways to improve on EWA Lite should attack the δ specification first.

when ϕ is low, the strategic environment is changing and information about past foregone payoffs is not likely to be a good guide to future choices. But then why should received payoffs be reinforced *relatively* more strongly than foregone payoffs (by a weight of one rather than low δ) when ϕ is low? There are two reasons. One is essentially econometric: When ϕ is low, then attractions from period $t-1$ are largely erased during the updating before period t. Reinforcing the chosen strategy from period t payoff with a weight of one partially "restores" information about what players are likely to do (since the erased lagged attractions and the previous choice which is strongly reinforced are likely to be correlated). The second reason is behavioral: Reinforcing chosen strategies more strongly than unchosen ones in low-ϕ environments models behavior of players who are especially likely to repeat what they did, like a "freezing" response to danger or "status quo bias", when the environment is changing.[11]

The second reason for the δ specification is that W is only greater than one in games which only have mixed equilibria. Why would a player want to have a lower δ (since $\delta = \frac{\phi}{W}$) in these games? A procedurally rational player may suspect that others are sophisticated and can anticipate, to some extent, what she will do by guessing how she learns. In a competitive game with a mixed equilibrium, being *too* responsive to foregone payoffs makes a player's choices too predictable, and makes her vulnerable to exploitation. Consider constant-sum games like matching pennies. If a player wins, the foregone payoffs from other strategies are zero so the value of δ makes no difference since it multiplied by the loss payoff (which is scaled to zero). But if she loses, then having a high δ implies you will switch to the (ex post) winning strategy, which makes you predictable and exploitable. If δ is low, however, then if you lose it is difficult to know

[11]Freezing is a response to danger which is nearly universal across species (including humans). This is presumably an adaptive response when predators are better at detecting movement than recognizing prey. Such a deep-seated response may lurk in the "old" or "animal" part of the human brain (the limbic system, which processes emotion and communicates with the prefrontal cortex that controls action; see Joseph LeDoux, 1996). Status quo bias refers to an exaggerated preference for the choice one has made in the past, even if the choice is assigned randomly (e.g., William Samuelson and Richard Zeckhauser, 1988; Kahneman, Knetsch, and Thaler, 1991). Experiments show that status quo bias is stronger when there are more sensible options available to switch to (loosely corresponding to W). This number-of-option effect can be captured in a model like ours by putting more reinforcement on the previous, status quo, choice, and putting reinforcement on alternative strategies which declines with W, precisely as in our model.

where you will switch to, which is actually a strategic advantage (and also characteristic of the data).

Third, and most decisively, the specification $\delta_i(t) = \phi_i(t)/W$ simply fits and predicts better than many others we tried (and than specifications in familiar reinforcement and belief models). The key reason is that in estimates from many games (see Camerer, Hsia, and Ho, 2000), $\hat{\delta}$ is generally between .5 and 1 in games with pure-strategy equilibria but is much lower (often zero) in games with mixed equilibria. Dividing ϕ by W pushes δ in the right direction– pushing it close to zero in games with mixed equilibrium– and enabling the $\delta_i(t)$ function to predict better across games than a fixed cross-game estimate of δ in the EWA model or in the prominent special cases.

2.3 The exploitation parameter $\kappa_i(t)$

The parameter κ controls the growth rate of attractions. When $\kappa = 0$ attractions are weighted averages of lagged attractions and ($\delta-$weighted) payoffs, so (if initial attractions are scaled to payoffs) the attractions are bounded by payoffs. If $\kappa = 0$ attractions cannot grow too far apart. Fixing λ across periods, this means it is difficult to predict very sharp convergence in later periods (as we sometimes see in the lab). That's because using the logit probability function, the degree of sharpness or convergence in probability (i.e., the difference between the highest and lowest choice probabilities) depends only on the *difference* in attractions, which is multiplied by sensitivity parameter λ. When attractions are bounded by payoffs, attractions cannot grow too far apart so the $\lambda-$weighted differences cannot be too large. (This could be remedied by choosing a higher λ but that predicts behavior which is sharp early one, contrary to observations.) When $\kappa = 1$, however, attractions are (decayed) cumulations of previous (weighted) payoffs. Then attractions can grow larger and larger– they can be multiples of payoffs– and consequently, choice probabilities can grow further apart.

Psychologically, κ can be interpreted as the extent to which players "explore" by choosing different strategies, relative to how quickly they "exploit" what they have learned by switching to a constant choice of the strategy which has performed the best in the past.[12]. Players with low κ are constantly exploring– they just keep track of average (δ-weighted) payoffs. When players "exploit" they commit to a strategy, even if its average previous payoff is not much larger than the average payoffs of other strategies. One way to model this is to let κ move toward one as players shift toward exploiting what they have learned. If payoffs are positive, a higher κ means players are basically rewarding a strategy they choose a lot, simply for being chosen (assuming $\delta < 1$). This is one way of characterizing lock-in empirically.

This line of argument suggests using variation in how frequently a player uses different strategies to track when they explore and when they exploit. We use the player's past behavior to tell us whether they explore or exploit and when they switch. The degree of exploration versus exploitation can be measured by the spread in probability of a player's observed choices. A standard measure of spread is the Gini coefficient, typically used to measure income inequality. We use the Gini coefficient too, where choice proportion is akin to income: When a player is exploring, the probabilistic 'income' will be spread to many strategies, and the Gini will be low. When a player has locked into one strategy, all her probability is allocated to that one and the Gini will be high (close to one).[13]

To calculate the Gini coefficient for subject i, first rank strategies from most-probable to least-probable (using observed choice frequencies). Denote the rank-ordered choice proportions of these strategies by $f_i^{(1)}(t)$ to $f_i^{(m_i)}(t)$. Then plot a cumulative probability distribution which measures the total probability of the strategies used as frequently as j or less frequently, $C_i(j,t) = \sum_{k=1}^{j} f_i^{(k)}(t)$. This calculation gives j points; use linear interpolation to create a piecewise-linear function connecting the points. The Gini coefficient is then the area between the identity line and the interpolated function passing through the $C_i(j,t)$ points, normalized so that Gini coefficients range from zero (when all strategies are played equally often) to 1 (when one strategy is played all the time).

The normalized Gini coefficient on strategy frequencies is then:

[12]The exploration- exploitation tradeoff is studied formally in the multi-armed bandit literature (John Gittins, 1989), and also of interest to computer scientists designing machines to learn, see Richard Sutton and Andrew Barto, 1998

[13]We also tried the sum of squared probabilities, a Herfindahl index often used to measure industrial concentration. This number is usually too low to fit well.

$$\kappa_i(t) = 1 - 2 \cdot \{\sum_{k=1}^{m_i} f_i^{(k)}(t) \cdot \frac{m_i - k}{m_i - 1}\} \tag{2.2}$$

where $f_i^k(t)$ are ranked from the lowest to the highest.[14] This κ function reflects the following thought process: A player tracks her actual choice frequencies. When the spread is low, the player is still exploring and wants to keep attractions from cumulating, so she chooses a low κ so that attractions continue to be averages. However, as she learns and chooses one strategy more often, she begins to exploit what she has learned. Exploitation requires a way of guaranteeing that the most frequently-chosen strategies get chosen more and more often. One way to do this is to let attractions cumulate, so that frequently-chosen strategy attractions will grow larger and larger simply because they are chosen more often (assuming $\delta < 1$). Letting κ be a function of strategy "concentration" is one way to do this.

Using cumulation to capture exploitation of high-payoff strategies is related to other ideas. One may be familiar to economists– Polya urns, which have been used to explain economies from increasing returns (Arthur, 1989) A Polya urn starts with a distribution of balls (e.g., some red and some black). When a red ball is drawn, it is replaced, along with *another* red ball. Draws therefore generate a payoff *and* increase the chance that the same payoff will occur again. This is a simple model of increasing returns or learning-by-doing (drawing a red makes red more likely) with interesting mathematical properties.

The Gini coefficient captures a similar process. If one strategy is chosen often that leads to large κ, which means that strategy payoffs cumulate. When $\delta < 1$ (as is common), cumulation favors chosen strategies; so strategies which are chosen often get chosen more often in the future, as in the Polya urn.

[14] For instance, in the median action game, suppose the relative choice frequencies for player i up to period t for actions 1-7 are 0, .0, .2, .4, .3, .0, and .1 respectively. Then we have $f_i^{(1)}(t) = f_i^{(2)}(t) = f_i^{(3)}(t) = 0.0$, $f_i^{(4)}(t) = 0.1$, $f_i^{(5)}(t) = 0.2$, $f_i^{(6)}(t) = 0.3$, and $f_i^{(7)}(t) = 0.4$. Consequently, we have: $\kappa_i(t) = 1 - 2 \cdot \{0 + 0 + 0 + 0.1 \cdot \frac{7-4}{7-1} + 0.2 \cdot \frac{7-5}{7-1} + 0.3 \cdot \frac{7-6}{7-1} + 0\}$, which is $\frac{2}{3}$.

2.4 Interpretation

The EWA Lite parameter functions are not grounded in principles familiar to game theorists. But the principles which *are* familiar, especially Bayesian updating, lead to models which EWA Lite strives (successfully) to outpredict. Any alternative model is necessarily different.

However, the EWA Lite parameter functions can be thought of as procedurally rational (in Herbert Simon's language) because they are precise and are designed to accomplish a specific goal: Namely, to predict and perform well in a wide range of games. One can imagine a truly optimal learning rule which maximizes expected payoffs across a wide range of games. However, we conjecture that such a rule would look more like EWA Lite than like other familiar rules. For example, fictitious play has good long-run properties in some environments but will not respond rapidly enough to changes in an environment. Cournot best-response changes too quickly in games with mixed equilibria. Weighted fictitious play is flexible enough to do well in both stationary environments and mixed games, but how does one pick the right weights? EWA Lite does so automatically based on what is observed. Rather than derive a globally rational approach from axioms, our approach is like work in machine learning, which tries to develop robust heuristic algorithms which learn effectively in a wide variety of low-information environments (see Sutton and Barto 1998). Good learning rules are not provably optimal but perform well on tricky test cases and lifelike problems like those which good computerized robots could perform (navigating around obstacles, hill-climbing on rugged landscapes, difficult pattern recognition, and so forth).

EWA Lite has three advantages. It is easy to use because it has only one free parameter to be estimated (λ).[15] The use of simple fictitious play and reinforcement theories in empirical analysis are often justified on the grounds of parsimony and how easily they

[15]It is conceivable that λ could also be specified ex ante but doing so will be difficult. The problem is that comparing values across games requires a standard unit of payoffs (we use dollars). However, changes in experimental currency which keep money earnings constant are likely to produce different behavior and require different values of λ (see Richard McKelvey, Thomas Palfrey, and Roberto Weber, forthcoming). Furthermore, the model implicitly assumes that differences in strategy attraction calibrated in money terms drive differences in choice probability but other framing effects may matter. For example, if players are sensitive to percentage differences in payoffs rather than absolute differences, then using a fixed λ

can be implemented. By those criteria, EWA Lite should be at least as attractive and will usually fit as well or better.

A second advantage is that parameters in EWA Lite naturally vary across time and people (as well as across games). In principle this might capture individual differences if they arise from experience. For example, some experiments on games with mixed-strategy equilibria allow subjects to explicitly choose randomized strategies (see Camerer, 2001, chapter 2). In these experiments, some subjects play pure strategies and some play mixtures. This difference in individual play is easily expressed by our $\kappa_i(t)$ function, which will tend toward one for purists and toward zero for mixers.

Because parameters can vary across time, a third advantage is that EWA Lite can mimic a very reduced form of "rule learning". Recall that different EWA parameter configurations correspond to specialized rules (such as cumulative choice reinforcement, fictitious play, or Cournot best-response dynamics). If parameters change throughout the game, those changes are like rule switching or rule learning, in which the rules players use change due to experience (as in Dale Stahl, 1996, 1999, forthcoming, and Tim Salmon, 1999). For example, if ϕ rises over time from 0 to 1, players are effectively switching from Cournot-type dynamics to fictitious play. If δ rises from 0 to 1, players are switching from reinforcement to belief learning. Rule learning is, of course, more general than the range of learning permissible in EWA Lite, and is an important competitor among those models which have many more parameters.

One way to think about EWA Lite is that it repairs weaknesses in reinforcement and belief learning and therefore proves more robust across games. Reinforcement learning falls short on two grounds: It assumes incorrectly that players only use information about the payoffs they received, even when they know foregone payoffs. Choice reinforcement sometimes underpredicts the rate of learning in games with large strategy spaces in which players initially choose strategies in one part of the space, then switch to strategies in an entirely different part of the space (Camerer and Ho, 1998; and see below for continental

across games will not explain what they choose (they will act like they use a lower λ when a positive constant is added to payoffs). John Pratt, David Wise, and Zeckhauser (1979) show this effect using field data on price dispersion across product categories. If all these effects are eventually understood a theory of how λ varies across games could be developed but such a goal is ambitious and beyond the scope of this paper.

divide games) because strategies which are chosen late in the game are rarely picked early on so they are not reinforced. Reinforcement also underpredicts movement in n-player games where most earn no profits in a period and get no reinforcement (such as auctions or "winner-take-all" labor tournaments). This is evident below in travellers' dilemma games.[16]

Belief learning has a different weakness. Belief models learn quickly in games with shifting support and many zero-payoff players, but give no ready explanation for why the decay rates on previous observations differ across games. For example, old observations are decayed more rapidly in the continental divide and beauty contest games than in games with mixed equilibria and patent race games. Confronted with a brand new game, belief learning theories have no built-in way of guessing whether Cournot-like responsive dynamics or standard fictitious play ($\phi = 1$) will predict best. EWA Lite can do better by predicting variation in ϕ as a function of initial conditions and game structure.

In the form we use, the EWA Lite model does require information about initial conditions (i.e, relative frequencies of first-period play) and information about the structure of the game– namely, the minimal support of Nash equilibria (W). As a practical matter, pinning down W boils down to guessing whether a game has a pure-strategy equilibrium, and whether it is symmetric or not, or has only mixed equilibria (and if so, how many strategies are used in the mixture). Even in field applications where the game is not controlled as in the lab, guessing whether W is one or much larger is probably not hard to do.

Finally, EWA Lite can be interpreted as a procedurally rational learning rule which attempts to adapt the nature of learning to features of the environment. A good example is the decay parameter ϕ. In games where other players are learning slowly, and outcomes are noisy (e.g., games with mixed equilibria), a good learning rule should have a value of ϕ close to one so that a large sample of past experience is taken. On the other hand,

[16]Of course, reinforcement learning can be speeded up if players reinforce payoffs relative to an aspiration level. But specifying aspiration levels requires two parameters– an initial apiration level and an adjustment rate. EWA generates aspiration-based reinforcement with no extra parameters: Strategies only increase in probability (holding their lagged attractions constant) if their $\delta-$weighted payoffs are above the average $\delta-$weighted payoff. Thus, the δ-weighted payoff is *is* an aspiration level, which evolves endogeneously over time without requiring any free parameters.

in games where learning is rapid, a player should use a low value of ϕ so that the effects of stale experiences are quickly discarded. An example is dominance-solvable p-beauty contests (Camerer, Ho and Chong, 2000). In our view, it is a modeling error to use the same value of ϕ for mixed games and for dominance-solvable games. The challenge in EWA Lite, however, is to create a "change-detection" function which allows ϕ to be low or high depending on the data a player sees.

3 EWA Lite predictions within and across seven games

In this section we compare in-sample fit and out-of-sample predictive accuracy of different learning models when parameters are freely estimated, and check whether EWA Lite functions can produce game-specific parameters similar to estimated values. We use seven games: Games with unique mixed strategy equilibrium (Mookerjhee and Sopher, 1997); R&D patent race games (Rapoport and Wilfred Amaldoss, 2000); a median-action order statistic coordination game with several players (Van Huyck, Battalio, and Richard Beil, 1990); a continental-divide coordination game, in which convergence behavior is extremely sensitive to initial conditions (Van Huyck, Joseph Cook, and Battalio, 1997); a coordination game about entry to two markets of different sizes (Amaldoss and Ho, 2001); dominance-solvable p-beauty contests (Ho, Camerer, and Keith Weigelt, 1998); and a traveler dilemma game (Monica Capra, Jacob Goeree, Rosario Gomez and Charles Holt, 1999). Table 1 summarizes features of these games and the data. Three of the games are described in detail below.[17] Since one of our goals is to see whether EWA Lite can explain cross-game variation in model parameters, we sample different classes of games. Sampling widely is also a good way to test robustness of any model of learning

[17]The other four games are: Mixed-equilibrium games studied by Mookerjhee and Sopher (1997) which have four or six strategies, one of which is weakly-dominated; the nine-player median-action game studied by Van Huyck et al. (1990), in which players choose integer strategies 1-7 and earn payoffs increasing linearly in the group median and decreasing linearly in the squared deviation from the median; dominance-solvable p-beauty contest games in which players choose numbers from 0 to 100 and the player closest to p times the average earns a fixed prize (for p equal to .7 or .9); and a coordination game in which n players simultaneously enter a large or small market and earn $2n$ (n) divided by the number of entrants if they enter the large (small) market.

or equilibrium.[18]

3.1 Estimation method

The estimation procedure for EWA Lite is sketched briefly here and detailed in Appendix. Consider a game where N subjects play T rounds. For a given player i, the likelihood function of observing a choice history of $\{s_i(1), s_i(2), \ldots, s_i(T-1), s_i(T)\}$ is given by:

$$\Pi_{t=1}^{T} P_i^{s_i(t)}(t). \tag{3.1}$$

The joint likelihood function L of observing all players' choice is given by

$$L(\lambda) = \Pi_i^N \{\Pi_{t=1}^{T} P_i^{s_i(t)}(t)\} \tag{3.2}$$

We "burn in" the model by choosing the initial attractions $A_i^j(0)$ (the same for all i) that correspond to choice probabilities that match the actual population frequency of choices in the first period (given the estimate of λ).[19] (When data on initial choices are unavailable some theory of initial play could be used instead.[20]) Details of the "burn-in" are given in Appendix 6.1. The initial parameter values are $\phi_i(0) = \kappa_i(0) = .5$ and $\delta_i(0) = \phi_i(0)/W$. These initial values are averaged with period-specific values determined by the functions, weighting the initial value by $\frac{1}{t}$ and the functional value by $\frac{t-1}{t}$.

[18]Another approach is to sample randomly within a class of games, although results are likely to be sensitive to which class of games is chosen.

[19]This is a small departure from some of our earlier work in which the $A_i^j(0)$ are estimated as free parameters. Estimation is infeasible in some of the games we study because there are many strategies (e.g., integer prices from 80 to 200) and we were reluctant to impose ad hoc functional forms to generate a parsimonious $A_i^j(0)$ distribution.

[20]A mixture of random behavior and "level-1" reasoning– best-responses to the belief that others will behave randomly– will generally be a good guess about what players would do in the first period (see Ernan Haruvy and Stahl, 1998). In the games we study, for example, level-1 behavior predicts choices of 35 in beauty contests, 7 in continental- divide games, 4 in median-action games, the large pot in entry-choice games, 5 and 4 in patent-race games for strong and weak players, and $200 - 2R$ in traveller's dilemma games. Combining these guesses with random initial behavior gives a good approximation to what players actually do in the first period.

Given the initial attractions and initial parameter values, attractions are updated using the EWA formula. EWA Lite parameters are then updated according to the functions above. Maximum likelihood estimation is used to find the best-fitting value of λ (and other parameters, for the other models) using data from the first 70% of the subjects. Then the value of λ is frozen and used to forecast behavior of the entire path of the remaining 30% of the subjects.[21] Payoffs were all converted to dollars (which is important for cross-game forecasting).

In addition to EWA Lite, we estimated the EWA model in Camerer and Ho (1999) and versions of belief-based (weighted fictitious play) and reinforcement models.[22] To put the models on a more even footing, we did *not* force the belief model to have initial attractions which are consistent with a common initial belief (as in our earlier work); we simply burned in the first-period data as for the other models.[23] We fit three versions of reinforcement ($\delta = 0$). Two versions included an experience weight $N(0)$ (which our 1999 paper did not) and fixed κ to be either zero or one (the latter is a simplified form of the model in Erev and Roth (1998)). A third version, which is quite different, is the two-parameter model used by Erev, Yoella Bereby-Meyer, and Roth (1999) and Roth et al. (2000). Their new approach sets $\phi = 1$ and $\kappa = 0$, updates only chosen strategies, uses logit probability instead of power, and divides attractions by a measure of payoff variability (see Appendix 6.3 for details). We report only results from this latest payoff-variability (PV) reinforcement model but performance of the earlier reinforcement models is similar. We also fit the one-parameter quantal response equilibrium (QRE) model (McKelvey and Palfrey, 1995, see Appendix 6.2 for details) as a static benchmark,

[21]This is another departure from our earlier work, in which we used the first 70% of the observations from each subject, then forecasted the last 30%. We also tried our earlier method, and a hybrid in which the holdout sample consisted of both later periods for some subjects, and the entire path for new subjects. The results from the two different methods are not interestingly different.

[22]For simplicity, we ignore two other interesting approaches to individual learning– rule learning or "learning to learn" (e.g., Stahl 1999; Salmon, 1999); and "direction learning" (Selten and Rolf Stoecker, 1986). See Camerer (2001, chapter 6) for more details.

[23]This switch helps belief models a lot in some games. For example, in the Mookerjhee-Sopher games with mixed equilibrium one strategy is only weakly dominated, and is rarely chosen. Most prior belief specifications will assign an expected payoff to that strategy which is only a little less than the expected payoffs of undominated strategies and given stochastic response, will overpredict how often the dominated strategy is chosen.

which is tougher competition than Nash equilibrium.

3.2 Model fit and predictive accuracy

The first question we answer is how well models fit and predict on a game-by-game basis (i.e., parameters are estimated separately for each game). To guard against overfitting we estimate parameters using 70% of the subjects (in-sample calibration) and use those estimates to predict choices by the remaining 30% (out-of-sample validation). For in-sample estimation we report both hit rates (the fraction of choices predicted to be most likely which are actually picked) and a Bayesian information criterion (BIC) which subtracts a penalty $\frac{k \cdot ln(NT)}{2}$ from the LL. (Note that the BIC imposes a stiffer penalty than other information criteria like Akaike.[24]) For out-of-sample validation we report hit rates and LL.

Table 2 shows the results. The best fits for each game and criterion are printed in bold; hit rates which are less than the best but are statistically indistinguishable (by the McNemar test) are italicized. Across games, EWA is better or as good as all other theories judging by hit rate, and fits better according to BIC or LL in four of seven games. EWA Lite also has higher or equal hit rates than other models in most games. Reinforcement with PV has the best BIC and LL in two games. Of the learning models, belief learning fits worst (it never has the best BIC or LL and is only best on hit rate in forecasting the median-action game). QRE fits worst of all, except in games with mixed equilibria where most models are about equally good.

The bottom line of each panel in Table 2, "pooled", shows results when a single set of common parameters is estimated for all games (except for game-specific λ). If EWA Lite is capturing parameter differences across games effectively, it should predict especially accurately, compared to other models, when games are pooled. Indeed, EWA Lite fits and predicts best by both criteria when data are pooled.

[24]In Jae Myung (2000) discusses model selection, including some recent measures which penalize theories for the flexibility of their functional form as well as for number of free parameters. He notes in an example that the squared deviation or squared error criterion (MSD, or MSE) is the measure which penalizes complex theories the *least* effectively. A very sensible measure, Bayesian model selection (BMS) reduces to the BIC when the modeller has a diffuse prior over parameter values.

A tough test of robustness is to estimate all parameters on six of the seven games and use those parameters to predict choices in the remaining game for all subjects, for each of the seven games. Cross-game prediction has been used by others but only within a narrow class of games (2x2 games with mixed equilibria, Erev and Roth, 1998; and 5x5 symmetric games, Stahl, forthcoming). Our results test whether fitting a model on a coordination game, say, can predict behavior in a game with mixed equilibrium. Table 3 reports results from this kind of cross-game prediction. EWA Lite has the highest hit rate in four games; EWA is highest in two other games. However, reinforcement with PV also predicts across games reasonably well; it has the best LL in three games. The biggest losers are belief models and QRE, which are usually much lower than the other models by any criterion.

The point of EWA Lite is to use only structural features of games (i.e., W) and players' experience to create parameter values which are close to the EWA estimates across games. Figure 2 shows how well EWA Lite functional values of δ and ϕ match the estimates from EWA. Each pair of connected points represents one of the seven games and the pooled estimates. Open (closed) circles are EWA estimates (EWA Lite functional values). If the two are close together within each game, and different across games, the chords connecting points should be short and sprinkled around the square. The chords are short in about half the games (most of the long-chord deviation between the function averages and estimate are on the ϕ dimension rather than δ). The correlation of the estimates and functional averages across games (excluding pooling) is .92 for δ and .78 for ϕ.[25] The pooled estimates only differ by .01 and .02, respectively. Details are reported in Table A.1 in Appendix (along with estimates for other models and standard errors).

In addition, Table A.2 in Appendix shows how much EWA Lite functional values vary across time periods and across people. Variation is usually not very large; the interquartile range is typically from zero to .10. An interesting exception is κ in the three games with mixed equilibria. The interquartile range for average κ within subject (i.e., averaged across periods) is .20 or more in these games. That means some subjects are roughly choosing pure strategies (κ near one) while others are mixing across all strategies (κ near zero), which shows the potential for the EWA Lite approach to detect

[25]The correspondence is much worse for κ, which is basically estimated to be either zero or one in EWA but only varies from around .4 to .8 in EWA Lite.

individual differences.

Next we will show predicted and relative frequencies for three games. Corresponding graphs for *all* games can be seen at http://www.fba.nus.edu.sg /depart/mk/fbacjk /ewalite/ewalite.html.[26] We chose these three games because each has interesting differences visible to the naked eye and each is representative of a different class– one has a unique mixed-strategy equilibrium, one has multiple Pareto-ranked pure equilibria, and one is dominance-solvable.

3.3 Games with unique mixed strategy equilibrium: Patent race

In the patent race game (Rapoport and Amaldoss, 2000), two players, one strong and one weak, are endowed with resources and compete in a patent race. The strong (weak) player has an endowment of 5 (4) and can invest an integer amount from zero to their endowments. Players invest simultaneously. They earn 10 minus their investment if their investment is strictly largest, and lose their investment if it is less than or equal to the other player's.

The game has an interesting strategic structure. The strong player can guarantee a payoff of five by investing the entire endowment (outspending the weak player), which strictly dominates investing zero. Eliminating the strong player's dominated (zero) strategy then makes investing one dominated for the weak player. Iterating in this way, both players can delete three strategies by iterated application of strict dominance. There is a unique mixed equilibrium in which strong (weak) players invest 5 (0) 60% of the time and play their other two (serially) undominated strategies 20% of the time.

Thirty six pairs of subjects played the game in a random matching protocol 160 times (with the role switched after 80 rounds); the 36 pairs are divided into 2 groups where random matching occurs within group. Choice frequencies do not change visibly across

[26]The website also has a GAUSS program readers can use to do their own estimation. Alternatively, readers who send us data, or a specification of their game, will receive estimates back for EWA Lite and any other models described in this paper, within a month.

time so we plot frequencies of transitions between period $t-1$ and period t strategies instead, for strong players, using the within-game estimation and pooling across all subjects. (Weak player results are similar.) Figures 3a-f show the empirical transition matrix and predicted transition frequencies for five models on strong players. The key features of the data are a lot of transitions from 5 to 5, almost 40%, and roughly equal numbers of transitions (about 5%) from 1 to 1, and from 1 to 5 or vice versa.

Two models are clearly inferior: QRE does not predict differences in transitions at all (it is a benchmark, not a learning theory); and the belief-based model predicts too few 5-to-5 and 1-to-1 transitions. (Table 2 confirms that belief learning fits relatively poorly here.) Where does belief learning go wrong? Since belief learning assumes full responsiveness to foregone payoffs, it will often predict that players should move away from chosen strategies which were winners, if other strategies would have been even better. Note that (as the equilibrium predicts) weak players abandon hope and invest zero about half the time. As a result, when strong players invest 5, half the time they earn a payoff of 5 but they could have earned more by investing less (because the weak player invested nothing). Belief models therefore predict more switching away from investing 5 than is evident in the data. Both EWA and reinforcement approaches can explain the infrequency of transitions by multiplying the higher foregone payoffs, in the case where the strong player invests 5 and the weak player invests nothing, by δ. A low value of δ (estimated to be .36 in EWA and .31 in EWA Lite) therefore characterizes the sluggishness in switching and avoids the predictive mistake inherent in belief learning.

3.4 Games with multiple pure strategy equilibria: Continental divide game

Van Huyck et al. (1997) studied a coordination game with multiple equilibria and extreme sensitivity to initial conditions, which we call the continental divide game (CDG). The payoffs in the game are shown in Table 4. Subjects play in cohorts of seven people. Subjects choose an integer from 1 to 14, and their payoff depends on their own choice and on the median choice of all seven players.

The payoff matrix is constructed so that there are two pure equilibria (at 3 and 12) which are Pareto-ranked (12 is the better one). Best responses to different medians are in bold. The best-response correspondence bifurcates in the middle: If the median starts at 7 virtually any sort of learning dynamics will lead players toward the equilibrium at 3. If the median starts at 8 or above, however, learning will eventually converge to an equilibrium of 12. Both equilibrium payoffs are shown in bold italics. The payoff at 3 is about half as much as at 12. This game captures the possibility of extreme sensitivity to initial conditions (or path-dependence).

Their experiment used 10 cohorts of seven subjects each, playing for 15 periods. At the end of each period subjects learned the median, and played again with the same group in a partner protocol. Payoffs were the amounts in the table, in pennies.

Figures 4a-f show empirical frequencies (pooling all subjects) and model predictions. The key features of the data are: Bifurcation over time from choices in the middle of the range (5-10) to the extremes, near the equilibria at 3 and 12; and late-period choices are more clustered around 12 than around 3. (Figure 4a disguises the extreme path-dependence: Groups which had first-period medians below (above) 7 *always* converged toward the low (high) equilibrium.) Notice also that strategies 1-4 are never chosen in early periods, but are frequently chosen in later periods; and notice that strategies 7-9 are frequently chosen in early periods but never chosen in later periods. A good model should be able to capture these subtle effects by "accelerating" low choices quickly (going from zero to frequent choices in a few periods) and "braking" midrange choices quickly (going from frequent choices to zero).

QRE fits poorly because it predicts no movement. Belief learning does not reproduce the asymmetry between sharp convergence to the high equilibrium and flatter frequencies around the low equilibrium. The reason why is diagnostic of a subtle weakness in belief learning. Note from Table 4 that the payoff gradients around the equilibria at 3 and 12 are exactly the same– choosing one number too high or low "costs" \$.02; choosing two numbers too high or low costs \$.08, and so forth. Since belief learning computes expected payoffs, and the logit rule means only differences in expected payoffs influence choice probability, the fact that the payoff gradients are the same means the spread of probability around the two equilibria must be the same; so belief learning will predict similar probability distributions around the high and low equilibria.

But how do the EWA and reinforcement models generate the asymmetry? The trick is a combination of $\delta < 1$ and cumulation (high κ). At the high equilibrium, the payoffs are larger and so the difference between the received payoff and δ times the foregone payoff will be larger than at the low equilibrium if $\delta < 1$, which explains the sharper convergence around 12.[27]

Reinforcement with PV fits reasonably well, except it predicts frequencies for choices 3-5 which are too low (10% instead of 15%) and it predicts substantial early choice of strategies 1-2 which declines over time. (The EWA models smear a little probability at 1-2, and grow over time, to avoid a large likelihood penalty from missing the rare choices of 1 and 2 which come in later periods.)

3.5 Games with dominance-solvable pure strategy equilibrium: Traveller's dilemma

Capra et al. (1999) studied a dominance-solvable "traveler's dilemma" (introduced by Kaushik Basu, 1984) in which two players must choose a number or 'claim' between 80 and 200. If the claims are equal, each player receives the amount claimed. If the claims are unequal, each of them receives the lower of the two claims. In addition, the person who makes the lower claim receives a bonus R and the person who makes the higher claim pays a penalty of R. Let these claims be x_1 and x_2 respectively. Formally, the payoff to each player i is defined as follows:

$$\pi_i(x_i, x_{-i}) = \begin{cases} x_i & \text{if } x_i = x_{-i} \\ x_i + R & \text{if } x_i < x_{-i} \\ x_{-i} - R & \text{if } x_i > x_{-i} \end{cases} \tag{3.3}$$

[27] Numerically, a player who chooses 3 when the median is 3 earns \$.60 and has a foregone payoff from 2 or 4 of \$.58 $\cdot\delta$. The corresponding figures for a player choosing 12 are \$1.12 and \$1.10 $\cdot\delta$. The differences in received and foregone payoffs around 12 and around 3 are the same when $\delta = 1$ but the difference around 12 grows larger as δ falls (for example, for the EWA Lite estimate $\hat{\delta} = .69$, the differences are \$.20 and \$.36 for 3 and 12.) Cumulating payoffs rather than averaging them contributes to explaining the difference by "blowing up" the expected payoff differences over time.

The game is like one of imperfect competition in which two sellers both sell products at the lowest price (due to consumer shopping or "meet-or-release" clauses) and the seller who names the lowest price earns a goodwill reward while the high-price sellers suffers a reputational loss. The Nash equilibrium predicts that the lowest possible claim of 80 will be chosen by both players. This prediction is also insensitive to R.

Their experiment used six groups of 9-12 subjects. The reward/penalty R was varied at 6 levels (5, 10, 20, 25, 50, 80). Each subject played 10 times (and played with a different R for five more rounds; we use only the first 10 rounds).[28]

Figures 5a-f show empirical frequencies and model fits for R=50. (Other values of R give roughly similar results although R=50 illustrates differences across models best.) A wide range of prices are named in the first round. Prices gradually fall, between 91-100 in rounds 3-5, 81-90 in rounds 5-6, and toward the equilibrium of 80 in later rounds.

QRE predicts a spike at the equilibrium of 80. As λ rises, the QRE equilibria move sharply from smearing probability throughout the price range to a sharp spike at the equilibrium; there is no intermediate value of λ which can explain the combination of initial dispersion and sharp convergence in later periods.

The belief-based model predicts the correct direction of convergence, but overpredicts numbers in the interval 81-90 and underpredicts choices of precisely 80. The problem is that the incentive in the travellers' dilemma is to undercut the other player's price by as little as possible. Players only choose 80 frequently in the last couple of periods; before those periods it pays to choose higher numbers. The belief model also estimates $\hat{\phi} = .85$ and does not allow payoffs to cumulate, so there is a large burden of historical information which keeps the model from reacting quickly to frequent choices of 80 which come late in the game. EWA models explain the sharp convergence in late periods by cumulating payoffs and estimating $\delta = .63$ (for EWA Lite). Consider two players who choose 80 and 90 when R=50. The player who chose 80 is reinforced by 130, and choosing the best response 89 is reinforced by $139 \cdot .63$, or 88.1, so choosing 80 is more strongly reinforced and she is likely to repeat that choice. In belief learning, the best response

[28] We did not use $R = 10$ with 9 subjects, where there is always one subject gone unmatched in each round, to avoid making ad hoc assumptions on the learning behavior of the unmatched subjects.

of 89 is reinforced by 139 so she is predicted to move *away* from 80, contrary to what happens.

The reinforcement model has a reasonable hit rate because the highest spikes in the graph often correspond with spikes in the data, but predicted learning is clearly more sluggish than in the data (i.e., the spikes are not high enough). Because effectively $\phi = 1$ and players are not predicted to move toward ex-post best responses, the model cannot explain why players learn to choose 80 so rapidly. (Reinforcement models with variable ϕ and no payoff variability term predict much better when R=50.)

4 Economic value of learning model

The criteria used to judge fit and predictive accuracy of models are purely statistical, or are roughly equivalent to familiar statistics. But economic applications of theories demand a financial measure of what good theories are worth.

In this section we measure the economic value of a learning theory. Camerer and Ho (2001) defined a theory's economic value as the increase in a subject's profit from substituting learning theory recommendations– best responses based on the theory's prediction about what others will do– for their actual choices. This definition treats a theory as similar to the advice service professionals (like consultants) sell, and measures its value by the difference in economic quality of the client's decisions with and without the advice.

To measure economic value, we use model parameters and a player's observed experience through period t to generate model predictions about what others will do in $t+1$. That prediction (a probability distribution over choices by others) produces a choice with the highest expected value. We then compare the profit from making that choice in $t+1$ (given what other players did in $t+1$) with profit from the target player's actual choice. Economic value is a good measure because it uses the full distribution of predictions about what other players are likely to do, *and* the economic impact of those possible choices.

We use two methods to estimate model parameters: Using in-game estimated parameters (see Appendix Table A.1); and using estimates from the six other games to compute economic value in the seventh game, for each of the seven games (as in Table 3). Using all the data from a game to give advice to subjects in that game is like advising a client who has a large sample of experience in a particular situation for the analyst to estimate parameters with. Using only data from other games is like advising a client in a new situation who has no direct experience for the analyst to use to estimate parameters. In practice, economic value would generally fall between the bounds of economic value computed these two ways. Note also that we do not control for the boomerang effect of a recommended choice on future behavior by others, but this effect will be small in most of the cases we study.[29]

Table 5 shows the overall economic value– the percentage improvement (or decline) in payoffs of subjects from following a model recommendation rather than their actual choices. Most models have positive economic value in most games. Value generated from in-game estimates is usually higher than when parameters are estimated from different games. Using either method of parameter estimation, either EWA Lite or EWA have the most economic value in five to seven of the seven games. Subject-by-subject analysis in Table 6 also shows that EWA and EWA Lite add value for a large majority of subjects[30], from 54% to 97% for out-of-game estimates (except pot games, where the figure is 48%).

Note that the percentage improvement is sometimes small because the difference between observed payoffs and ex-post optimal payoffs– the payoffs that a perfectly clairvoyant player would have earned– is often low, which puts an upper bound on absolute

[29]That is, if a player actually chose 9 in the continental divide game and a theory advised choosing 10, in future periods we continue to assume the subject chose 9. Note however that in the coordination and p-beauty contest games medians and averages are computed for groups of 7-9 players, so changing one player's choice will not affect overall behavior much. In the travellers' dilemma players are randomly rematched so changing one player's current choice may affect her current partner's future behavior, but she is unlikely to be rematched with same partner. The effect could be more substantial in mixed and patent race games, and in the pot games with small numbers of subjects. The obvious correction, which we are pursuing, is to run experiments in which one or more computerized subjects actually use a learning model to make choices, and compare their performance with that of actual subjects.

[30]In some games many estimates of economic value are zero because subjects make the same choices the learning theories recommend.

economic value. In the continental divide game, for example, perfect forecasts would have improved profits by 6.58% and EWA Lite improved profits (out-of-game) by 4.98%, so the EWA Lite forecast improve profits by about 80% of the largest possible improvement.

Belief learning also has positive economic value in almost all games. Reinforcement actually has negative economic value in one to three games. In continental divide games, reinforcement tends to underpredict how rapidly players move away from middle strategies. It therefore advises players to switch to low or high strategies less quickly than they actually do, which turns out to be bad advice. The beauty contest game is similar. Reinforcement predicts hardly any learning by other players, so it never advises players to chose numbers which are low enough.

5 Conclusion

Learning is important for economics. Equilibrium theories are useful because they suggest a possible limit point of a learning process and permit comparative static analysis. But if learning is slow, or the time path of behavior selects one equilibrium out of many, a precise theory of equilibration could prove useful.

In the last ten years, many theories of individual learning in games have been proposed and fit to data from laboratory games in which experimenters have good control over players' information and incentives. Some of these theories, particularly reinforcement learning and fictitious play or Cournot belief learning, are very simple (i.e., they have few free parameters which have to be specified or estimated from data). Simple theories have the advantage of parsimony but often miss important features of empirical data and, importantly, can be improved by adding features judiciously. Other theories are quite complex (e.g., Crawford, 1995; Stahl, 1999).

Because there are many theories, applied to different games using different scientific standards of proof or utility, some healthy controversy has emerged about which models are best for which purposes. Our earlier theory (Camerer and Ho, 1999), parametric EWA (EWA for short) takes a middle road by hybridizing features of reinforcement and belief learning, which necessarily adds parameters. Estimates across two dozen experimental data sets show that adding these features improves fit and predictive accuracy, but the parameter values which maximize fit are typically significantly different in different games. This finding raises the question of how to predict in advance which parameter values will fit best in a particular game.

The theory described in this paper, EWA Lite, replaces three parameters in the EWA learning models with three functions that change over time in response to experience. One function is a "motion detector" ϕ which goes up (limited by one) when behavior by other players is stable, and dips down (limited by zero) when there is surprising new behavior by others. When ϕ dips down, the effects of old experience (summarized in attractions which cumulate or average previous payoffs) is diminished by decaying the old attraction by a lot. The second function δ is simply ϕ divided by W, the minimal number of strategies in equilibrium. This function ties responsiveness to foregone payoffs to environmental stability, and also lowers δ in games with mixed equilibria ($W > 1$). The third function κ is an index of concentration of choices (a normalized Gini coefficient). This characterizes players who "explore" a lot (trying different strategies, yielding a low κ) and those who "exploit" by locking in to a single choice (high κ). EWA Lite is more parsimonious than most learning theories because it has only one free parameter– the response sensitivity λ.

We fit and predict data from seven different experimental games using EWA Lite, our previous parameterized EWA model, and three other models (belief learning, reinforcement with payoff variability, and quantal response equilibrium (QRE)). Note that QRE and EWA Lite both have one free parameter, reinforcement has two, belief learning has three, and EWA has five. We always report both in-sample fit (penalizing more complex theories using the Bayesian information criterion) and out-of-sample predictive accuracy to be sure that more complex models do not necessarily fit better.

There are three key results.

First, EWA Lite solves a potential problem with application of the EWA model across games because it endogenizes EWA parameters effectively. It fits and predicts about as accurately as EWA in all seven games; and it produces functional parameter values for δ and ϕ which track estimated values quite closely across games. Because EWA Lite generates sensible cross-game parameter variation automatically, it fits and predicts better than all other models when games are pooled and common parameters are estimated.

Second, we use a new criterion for judging the usefulness of theories, called economic value (introduced by Camerer and Ho, 2001). A theory's economic value is the incremental profit a subject would earn from following the theory's advice rather than making their own choices. Either EWA or EWA Lite add the most economic value (and add positive value for a majority of subjects), in five or seven of the seven games, depending on how parameters are estimated to provide advice. Reinforcement models actually have negative economic value in one to three games (depending on estimation).

What's the bottom line? Because we used many criteria and games, it is not surprising that no one theory is always best. The results are sensitive to which games are used, but not sensitive to performance criteria. In coordination and dominance-solvable games, either EWA or EWA Lite fit and predict best and add the most economic value. In mixed-strategy games reinforcement fits a little better than the EWA models by statistical criteria, and adds similar economic value. Belief models hardly ever fit best (EWA Lite is simpler and almost always fits better). And all learning theories fit reliably better than QRE.

A next step in this research is to apply EWA Lite to a wider set of games. We would also like to find some axiomatic underpinnings for the functions, which are admittedly ad hoc. Extending the ϕ function to exploit information about ordered strategies might prove useful. And since EWA Lite is so parsimonious, it is useful as a building block for extending learning theories to include sophistication (players anticipating that others are learning; see Stahl, 1999) and explain "teaching" behavior in repeated games (Camerer, Ho and Chong, 2000; David Cooper and John Kagel, 2001).

The theory is developed to fit experimental data, but the bigger scientific payoff will come from application to naturally-occurring situations. If learning is slow, a precise theory of economic equilibration is just as useful for predicting what happens in the economy as a theory of equilibrium. For example, institutions for matching medical residents and medical schools, and analogous matching in college sororities and college bowl games, developed over decades (Roth and Xiaolin Xing, 1994). Bidders in eBay auctions learn to bid late to hide their information about an object's common value (Patrick Bajari and Ali Hortacsu, 1999). Consumers learn over time what products they like (Ho and Chong, 1999). Learning in financial markets can generate excess volatility and returns predictability, which are otherwise anomalous in rational expectations models (Allan Timmerman, 1993). Thomas Sargent (1999) argues that learning by policymakers about expectational Phillips' curves and the public's perceptions of inflation explains macroeconomic behavior in the last couple of decades. Good theories of learning should be able to explain these patterns and help predict how new institutions will evolve, how rapidly bidders learn to wait, and which new products will succeed. Applying EWA Lite to field domains is therefore an important goal of future research.

References

[1] Amaldoss, Wilfred and Ho, Teck-Hua, "EWA Learning in Games with Different Group Sizes," Marketing Department Working Paper, The Wharton School, 2001.

[2] Arthur, Brian, "Competing Technologies, Increasing Returns, and Lock-in by Historical Events," *Economic Journal*, 1989, 99, pp. 116-131.

[3] Arthur, Brian, "Designing Economic Agents That Act Like Human Agents: A Behavioral Approach to Bounded Rationality," *American Economic Review*, 1991, 81(2), pp. 353-359.

[4] Bajari, Patrick and Hortacsu, Ali, "Winner's Curse, Reserve Prices and Endogenous Entry: Empirical Insights from Ebay Auctions," Working Paper, Department of Economics, Stanford University, 1999.

[5] Basu, Kaushik, "The Traveler's Dilemma: Paradoxes of Rationality in Game Theory," *American Economic Review*, May 1984, 84(2), pp. 391-395.

[6] Brown, George, "Iterative Solution of Games by Fictitious Play," in , ed., *Activity Analysis of Production and Allocation*, New York: John Wiley & Sons, 1951.

[7] Bush, Robert and Mosteller, Frederick, *Stochastic models for learning*, New York: Wiley, 1955.

[8] Camerer, Colin, *Behavioral Game Theory: Experiments on Strategic Interaction*, Princeton University Press, 2001.

[9] Camerer, Colin and Ho, Teck-Hua, "Experience-Weighted Attraction Learning in Coordination Games: Probability Rules, Heterogeneity and Time-Variation," *Journal of Mathematical Psychology*, 1998, 42, pp. 305-326.

[10] Camerer, Colin and Ho, Teck-Hua, "Experience Weighted Attraction Learning in Normal Form Games," *Econometrica*, 1999, 67, pp. 827-873.

[11] Camerer, Colin and Ho, Teck-Hua, "Strategic Learning and Teaching in Games," in S. Hoch and H. Kunreuther, eds., *Wharton on Decision Making*, New York: Wiley, 2001.

[12] Camerer, Colin, Ho-Teck-Hua, and Wang, Xin, "Individual Differences in EWA Learning with Partial Payoff Information," Marketing Department Working Paper 99-010, The Wharton School, 1999.

[13] Camerer, Colin, Ho, Teck-Hua and Chong, Juin-Kuan, "Sophisticated EWA Learning and Strategic Teaching in Repeated Games," Marketing Department Working Paper 00-005, The Wharton School, 2000.

[14] Camerer, Colin, Hsia, David and Ho, Teck-Hua, "EWA Learning in Bilateral Call Markets," Marketing Department Working Paper 00-021, The Wharton School, 2000.

[15] Capra, Monica, Goeree, Jacob, Gomez, Rosario and Holt, Charles, "Anomalous Behavior in a Traveler's Dilemma," *American Economic Review*, 89(3), June 1999, pp. 678-690.

[16] Chen, Yan and Khoroshilov, Yuri, "Learning under Limited Information," Working Paper, Department of Economics, University of Michigan, Ann Arbor, 2000.

[17] Cheung, Yin-Wong and Friedman, Daniel, "Individual Learning in Normal Form Games: Some Laboratory Results," *Games and Economic Behavior*, 1997, 19, pp. 46-76.

[18] Cooper, David and John Kagel. Learning and Transfer in Signalling Games. Department of Economics, Case Western Reserve University, 2001.

[19] Crawford, Vincent P., "Adaptive Dynamics in Coordination Games," *Econometrica*, 1995, 63, pp. 103-143.

[20] Crick, Francis. The Mad Pursuit, 1988.

[21] Cross, John, *A Theory of Adaptive Learning Economic Behavior*, New York: Cambridge University Press, 1983.

[22] Erev, Ido and Roth, Alvin E., "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria," *American Economic Review*, 1998, 88(4), pp. 848-81.

[23] Erev, Ido, Bereby-Meyer, Yoella and Roth, Alvin E., "The Effect of Adding a Constant to All Payoffs: Experimental Investigation, and a Reinforcement Learning Model with Self-Adjusting Speed of Learning," *Journal of Economic Behavior and Organization*, 1999, 39, pp. 111-128.

[24] Fudenberg, Drew and Levine, David K., *The Theory of Learning in Games*, Boston: MIT Press, 1998.

[25] Gittins, John, *Multi-armed Bandit Allocation Indices*, New York: Wiley, 1989.

[26] Haruvy, Ernan and Stahl, Dale O., " An Empirical Model of Equilibrium Selection in Symmetric Normal-form Games," University of Texas Department of Economics Working Paper, January 1998.

[27] Harley, Calvin, "Learning the Evolutionary Stable Strategies," *Journal of Theoretical Biology*, 1981, 89, pp. 611-633.

[28] Heller, Dana and Sarin, Rajiv, "Parametric Adaptive Learning," University of Chicago Working Paper, 2000.

[29] Ho, Teck-Hua and Chong, Juin-Kuan, "A Parsimonious Model of SKU Choice," Marketing Department Working Paper 99-020, The Wharton School, 1999.

[30] Ho, Teck-Hua, Camerer, Colin and Weigelt, Keith, "Iterated Dominance and Iterated Best Response in Experimental "p-Beauty Contests"," *American Economic Review*, 1998, 88, pp. 947-969.

[31] Hopkins, Edward, "Two Competing Models of how People Learn in Games," University of Edinburgh and Pittsburgh Working Paper, 1999.

[32] Kahneman, Daniel, Knetsch, Jack L. and Thaler, Richard H.,"The Endowment Effect, Loss Aversion, and Status Quo Bias: Anomalies," *Journal of Economic Perspectives*, Winter 1991, 5(1), p.193-206.

[33] LeDoux, Joseph, *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*, New York : Simon & Schuster, 1996.

[34] McAllister, Patrick H., "Adaptive Approaches to Stohcastic Programming," *Annals of Operations Research*, 1991, 30, pp. 45-62.

[35] McKelvey, Richard and Palfrey, Thomas, "Quantal Response Equilibria for Normal Form Games," Games and Economic Behavior, 1995, 10, 6-38.

[36] McKelvey, Richard, Palfrey, Thomas and Weber, Roberto, "The Effects of Payoff Magnitude and Heterogeneity on Behavior in 2x2 Games with Unique Mixed Strategy Equilibria," *Journal of Economic Behavior and Organization*, 2000, (forthcoming).

[37] Mookerjhee, Dilip, and Sopher, Barry, "Learning Behavior in an Experimental Matching Pennies Game," *Games and Economic Behavior*, 1994, 7, pp. 62-91.

[38] Mookerjhee, Dilip, and Sopher, Barry, "Learning and Decision Costs in Experimental Constant-sum Games," *Games and Economic Behavior*, 1997, 19, pp. 97-132.

[39] Myung, In Jae, "The importance of complexity inn model selection," *Journal of Mathematical Psychology*, 2000, 44, pp. 190-204.

[40] Pratt, John, Wise, David and Zeckhauser, Richard, "Price Differences in Almost Competitive Markets," *Quarterly Journal of Economics,*, May 1979, 93(2), pp. 189-211.

[41] Rapoport, Amnon and Erev, Ido, "Coordination, "magic", and reinforcement learning in a market entry game," *Games and Economic Behavior*, 1998, 23, pp. 146-175.

[42] Rapoport, Amnon and Amaldoss, Wilfred, “Mixed Strategies and Iterative Elimination of Strongly Dominated Strategies: An Experimental Investigation of States of Knowledge,” *Journal of Economic Behavior and Organization*, 2000, 42, pp. 483-521.

[43] Robinson, Julia, “An Iterative Method of Solving a Game,” *Annals of Mathematics*, 1951, 54(2), pp. 296-301.

[44] Roth, Alvin E. and Xing, Xiaolin, “Jumping the Gun: Imperfections and Institutions Related to the Timing of Market Transactions,” *American Economic Review*, 84, September, 1994, 992-1044.

[45] Roth, Alvin E. and Erev, Ido, “Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term,” *Games and Economic Behavior*, 8(1), 1995, pp. 164-212.

[46] Roth, Alvin, Barron, Greg, Erev, Ido and Slonim, Robert, “Equilibrium and Learning in Economic Environments: the Predictive Value of Approximations,” Harvard University Working Paper, 2000.

[47] Salmon, Tim, “Evidence for Learning to Learn Behavior in Normal Form Games,” Caltech Working Paper, 1999.

[48] Samuelson, William and Zeckhauser, Richard, ”Status Quo Bias in Decision Making,” *Journal of Risk and Uncertainty*, March 1988, 1, pp. 7-59.

[49] Sargent, Thomas, *The Conquest of American Inflation*, Princeton: Princeton University Press, 1999.

[50] Selten, Reinhard and Stoecker, Rolf, “End Behavior in Sequences of Finite Prisoner’s Dilemma Supergames: A Learning Theory Approach,” *Journal of Economic Behavior and Organization*, 1986, 7, pp. 47-70.

[51] Stahl, Dale O, “Boundedly Rational Rule Learning in a Guessing Game,” *Games and Economic Behavior*, 1996, 16, pp. 303-330.

[52] Stahl, Dale O., “Sophisticated Learning and Learning Sophistication,” University of Texas at Austin Working Paper, 1999.

[53] Stahl, Dale O, "Local Rule Learning: Theory and Evidence," *Games and Economic Behavior*, 2000 (forthcoming).

[54] Sutton, Richard and Barto, Andrew, *Reinforcement Learning: An Introduction*, Boston: MIT Press, 1998.

[55] Timmerman, Allan G, "How Learning in Financial Markets Generates Excess Volatility and Predictability in Stock Prices.' *Quarterly Journal of Economics*, November 1993, 108, pp. 1135-1145.

[56] Van Huyck, John, Battalio, Raymond and Beil, Richard, "Tacit Cooperation Games, Strategic Uncertainty, and Coordination Failure," *The American Economic Review*, 1990, 80, pp. 234-248.

[57] Van Huyck, John, Battalio, Raymond and Raskin, Frederick, "Selection Dynamics and Adaptive Behavior Without Much Information," Working Paper, Department of Economics, Texas A & M Uuiversity, 1996.

[58] Van Huyck, John, Cook, Joseph and Battalio, Raymond, "Adaptive Behavior and Coordination Failure," *Journal of Economic Behavior and Organization*, 1997, 32, pp. 483-503.

6 Appendix

6.1 Calculating the Initial Attractions

We "burn in" the initial attractions $A^j(0)$, $\forall j$, by using the actual observed frequency of choices by all subjects in the first period. The same initial attractions are used for all subjects, except for games with more than 2 players such as continental divide, median action and p-beauty contest. For the exceptions, each group of players has a different initial attractions based on the group's observed frequencies. For a particular response sensitivity λ, the initial attractions are chosen so that the predicted probabilities of choices match the actual relative frequencies of choices.

Denote the empirically observed frequency of strategy j in the first period by f^j. Then initial attractions are recovered from the equations

$$\frac{e^{\lambda \cdot A^j(0)}}{\sum_k e^{\lambda \cdot A^k(0)}} = f^j, j = 1, \ldots, m. \tag{6.1}$$

(This is equivalent to choosing initial attractions to maximize the likelihood of the first-period data, separately from the rest of the data, for a value of λ derived from the overall likelihood-maximization.) Some algebra shows that the initial attractions can be solved for, as a function of λ, by

$$A^j(0) - A^k(0) = \frac{1}{\lambda} ln(f^j) - \frac{1}{\lambda} ln(f^k), \quad j, k = 1, \ldots, m \tag{6.2}$$

We fix the initial attraction of the strategy j with the lowest frequency $A^j(0)$ to a constant value for identification. Frequently, the lowest frequency is zero. We circumvent this problem by adding a constant $\frac{W}{m}$ to all frequencies and renormalizing them.

$$\tilde{f}^j = \frac{f^j + \frac{W}{m}}{\sum_k f^k + \frac{W}{m}}, \quad j = 1, \ldots, m \tag{6.3}$$

$\frac{W}{m}$ is chosen to reflect the relative proportion of equilibrium points with respect to number of strategies. With $\tilde{f}^j$ in place of f^j in (6.2), we then solve for the other attractions as a function of λ and the modified frequencies $\tilde{f}^j$.

To ensure no model obtains any unfair advantage from the burn in procedure, we use (6.1) as the first period prediction for all models.

Since the model specifications for EWA Lite, EWA and Belief-based learning have been discussed in the text, we only provide the model specifications for the remaining models, Quantal Response and Reinforcement with Payoff Variability, below.

6.2 Quantal Response Model

The updating rule and predicted probability are given as follows:

$$A_i^j(t) = \sum_{k=1}^{m_{-i}} P_{-i}^k(t+1) \cdot \pi_i(s_i^j, s_{-i}^k(t)) \tag{6.4}$$

$$P_i^j(t+1) = \frac{e^{\lambda \cdot A_i^j(t)}}{\sum_{k=1}^{m_i} e^{\lambda \cdot A_i^k(t)}} \tag{6.5}$$

As evident from (6.5), the predicted probability is a function of other player(s) predicted probabilities. For a given sensitivity parameter λ, predicted probabilities are derived from solving N nonlinear simultaneous equations. We solve the nonlinear simultaneous equations numerically by iterative substitutions until we converge to a set of consistent predicted probabilities.

6.3 Reinforcement Model with Payoff Variability

The updating rule is:

$$A_i^j(t) = \frac{(N(0) + C_{ij}(t) - 1) \cdot A_i^j(t-1) + I(s_i^j, s_i(t)) \cdot \pi_i(s_i^j, s_{-i}(t))}{N(0) + C_{ij}(t)} \tag{6.6}$$

where $C_{ij}(t)$ (with $C_{ij}(0) = 0, \ \forall i, j$) is updated as follows:

$$C_{ij}(t) = \begin{cases} C_{ij}(t-1) + 1 & \text{if } j \text{ is chosen in } t \\ C_{ij}(t-1) & \text{if } j \text{ is not chosen in } t \end{cases} \tag{6.7}$$

In addition, λ is replaced by $\frac{\lambda}{S_i(t)}$ where

$$S_i(t) = \frac{(t - 1 + m \cdot N(0))S_i(t-1) + |\overline{\pi_i(t-1)} - \pi_i(s_i(t), s_{-i}(t))|}{t + m \cdot N(0)} \tag{6.8}$$

where m is the number of strategies and $A_i(t)$ is updated as follows:

$$\overline{\pi_i(t)} = \frac{(t - 1 + m \cdot N(0))\overline{\pi_i(t-1)} + \pi_i(s_i(t), s_{-i}(t))}{t + m \cdot N(0)} \tag{6.9}$$

where $\overline{\pi_i(0)}$ is the expected payoff given random choice. Instead of assuming random choices by other players in the computation of $\overline{\pi_i(0)}$, we use empirical distribution of other players' first period choices to increase the potency of the model. This also ensures that the Reinforcement Model with Payoff Variability is placed on the same footing as other models where first period is used to burn in initial attractions. $S_i(0)$ is the expected absolute difference between payoff from each strategy and $\overline{\pi_i(0)}$.

6.4 Parameter Estimates and Functional Values

Table A.1 gives the parameter estimates and their standard errors of all learning models. Note that the standard errors are small, suggesting that these parameters are statistically different from zero. Table A.2 shows the inter-quartile ranges of EWA Lite functional values across time and subjects. Except for κ in some games, the ranges are relatively small.

Table 1: A Description of the Seven Games Used in the Estimation of Various Learning Models

Game	Number of Players	Number of Strategies	Number of Pure Strategy Equilibria	Number of Subjects	Number of Rounds	Matching Protocol	Experimental Treatment	Description of Games
Mixed Strategies Mookerjhee and Sopher (1997)	2	4,6	0	80	40	Fixed	Stake Size	A constant-sum game with unique mixed strategy equilibrium.
Patent Race Rapoport and Amaldoss (2000)	2	5,6	0	36	80	Random	Strong vs Weak	Strong (weak) player invests between 0 and 5 (0 and 4) and the higher investment wins a fixed prize.
Continental Divide Van Huyck et al. (1997)	7	14	2	70	15	Fixed	None	A coordination game with two pure strategy equilibria
Median Action Van Huyck et al. (1990)	9	7	7	54	10	Fixed	None	A order-statistic game with individual payoff decreases in the distance between individual choice and the median
Pot Games Amaldoss and Ho (2001)	3,6,9,18	2	1	84	25 (manual) 28 (computer)	Fixed	Number of Players	An entry game where players must decide which of the two ponds of sizes 2n and n they wish to enter. Payoff is the ratio of the pond size and number of entries.
Traveller's Dilemma Capra et al. (1999)	2	121[1]	1	52	10	Random	Penalty Size	Players choose claims between 80 and 200. Both players get lower claim but the high-claim player pays a penalty to the low-claim player.
p-Beauty Contest Ho et al. (1998)	7	101	1	196	10	Fixed	Experienced vs. Inexperienced	Players simultaneously choose a number from 0 to 100 and the winner whose number is closet to p (<1) times the group average

Note 1: Continuous strategies of 80 to 200 are discretized to 121 integer strategies

Table 2: Model Fit and Prediction (% Hit Rate, BIC and Log Likelihood)

In-sample Calibration

	Sample Size[5]	EWA Lite		EWA		Belief-based		Reinforcement with PV		QRE	
		%Hit[1,4]	BIC[2]	%Hit	BIC	%Hit	BIC	%Hit	BIC	%Hit	BIC
Mixed Strategies	2240	40%	-3192	**41%**	-3074	38%	-3129	*41%*	**-3051**	31%	-3340
Patent Race	4000	*62%*	-4442	*61%*	-4411	53%	-5506	**62%**	**-4367**	38%	-6682
Continental Divide	735	*50%*	-1081	**51%**	**-1062**	30%	-1289	47%	-1293	6%	-1888
Median Action	380	69%	-313	**75%**	**-272**	*74%*	-348	69%	-343	50%	-541
Pot Games	1478	**68%**	**-905**	*67%*	-937	65%	-982	*66%*	-907	62%	-1018
Traveller's Dilemma	360	**52%**	-889	*50%*	**-871**	31%	-1115	46%	-1070	25%	-1764
p-Beauty Contest	1380	**13%**	-4567	*13%*	**-4539**	*13%*	-4578	10%	-5736	3%	-5844
Pooled [3]	10573	**51%**	**-15388**	48%	-15906	40%	-17960	44%	-20182	33%	-21055

Out-of-sample Validation

	Sample Size	EWA Lite		EWA		Belief-based		Reinforcement with PV		QRE	
		%Hit	LL	%Hit	LL	%Hit	LL	%Hit	LL	%Hit	LL
Mixed Strategies	960	**36%**	**-1382**	*36%*	-1387	*34%*	-1405	*33%*	-1392	*35%*	-1400
Patent Race	1760	*65%*	-1897	**65%**	-1878	53%	-2279	*65%*	**-1864**	40%	-2914
Continental Divide	315	*47%*	-470	**47%**	**-460**	25%	-565	44%	-573	6%	-805
Median Action	160	74%	-104	*79%*	**-83**	**82%**	-95	74%	-105	49%	-187
Pot Games	739	**70%**	-436	**70%**	-437	66%	-471	*70%*	**-432**	65%	-505
Traveller's Dilemma	160	**46%**	-445	*43%*	**-443**	36%	-465	*41%*	-561	27%	-720
p-Beauty Contest	580	**8%**	-2119	*6%*	**-2042**	*7%*	-2051	*7%*	-2494	3%	-2502
Pooled	4674	**51%**	**-6852**	49%	-7100	40%	-7935	46%	-9128	36%	-9037

Note 1: Number of hits counts the occasions when prob(chosen strategy) = maximum (predicted probabilities). Each count is adjusted by number of strategies sharing the maximum.

Note 2: BIC (Bayesian Information Criterion) is given by LL - (k/2)*log(N*T) where k is the number of parameters, N is the number of subjects and T is the number of periods.

Note 3: A common set of parameters, except game-specific lambda, is estimated for all games. Each games is given equal weight in LL estimation.

Note 4: Entries in bold denote the best measures for each game. In case of hit rate, multiple models might share the top rank when differences in hit rates of these models are not statistically significant by McNemar test (Chi-sq at 5%); these entries are italicized.

Note 5: Calibrated on all observations for 70% of the subjects instead of 70% observations of all subjects as in Camerer and Ho (1999).

Table 3: Model Robustness (Out-of-sample Prediction for New Game)

		Out-of-sample Prediction using Out-of-game Estimates[1]									
	Sample	EWA Lite		EWA		Belief-based		Reinforcement with PV		QRE	
	Size	%Hit	LL	%Hit	LL	%Hit	LL	%Hit	LL	%Hit	LL
Mixed Strategies	3200	**39%**	**-4662**	34%	-4867	35%	-4832	*38%*	-4697	31%	-5049
Patent Race	5760	*62%*	-8009	59%	-9296	55%	-9112	**63%**	**-6588**	39%	-9745
Continental Divide	1050	48%	-1741	**50%**	**-1635**	27%	-2147	32%	-2403	6%	-2695
Median Action	540	69%	-523	**74%**	-491	60%	-711	70%	**-479**	50%	-990
Pot Games	2217	**68%**	-3976	67%	-5084	65%	-3474	67%	**-1387**	63%	-1491
Traveller's Dilemma	520	**50%**	**-1825**	46%	-1874	33%	-1933	36%	-1841	21%	-2325
p-Beauty Contest	1960	**12%**	**-6681**	10%	-6819	8%	-7715	9%	-11361	3%	-8342

Note 1:Prediction for a game is made using out-of-game estimates derived from pooling the other 6 games.

Table 4: Payoffs in 'continental divide' experiment, Van Huyck et al. (1997)

	Median Choice													
choice	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	45	49	52	55	56	55	46	-59	-88	-105	-117	-127	-135	- 142
2	**48**	53	58	62	65	66	61	-27	-52	-67	-77	-86	-92	-98
3	**48**	**54**	***60***	**66**	70	74	72	1	- 20	-32	-41	-48	-53	-58
4	43	51	58	65	**71**	**77**	80	26	8	-2	-9	-14	-19	-22
5	35	44	52	60	69	**77**	**83**	46	32	25	19	15	12	10
6	23	33	42	52	62	72	82	62	53	47	43	41	39	38
7	7	18	28	40	51	64	78	75	69	66	64	63	62	62
8	-13	-1	11	23	37	51	69	83	81	80	80	80	81	82
9	-37	-24	-11	3	18	35	57	88	89	91	92	94	96	98
10	-65	-51	-37	-21	-4	15	40	**89**	**94**	98	101	104	107	110
11	-97	-82	-66	-49	-31	-9	20	85	**94**	**100**	105	110	114	119
12	-133	-117	-100	-82	-61	-37	-5	78	91	99	**106**	***112***	**118**	**123**
13	-173	-156	-137	-118	-96	-69	-33	67	83	94	103	110	117	**123**
14	-217	-198	-179	-158	-134	-105	-65	52	72	85	95	104	112	120

Table 5: Economic Values (Performance of a Bionic Subject) with In-game and Out-of-game Estimates

Total Payoff and Percentage Improvement for Bionic Subjects using Out-of-game Estimates[1]

	Observed Payoff	Ex-post %Improve	EWA Lite %Improve	EWA %Improve	Belief-based %Improve	Reinforcement with PV %Improve	QRE %Improve
Mixed Strategies	334	100.0%	**7.5%**	3.0%	1.1%	5.8%	-1.8%
Patent Race	467	44.2%	1.7%	1.2%	1.3%	**2.9%**	1.2%
Continental Divide[2]	837	6.6%	5.0%	**5.2%**	4.5%	-9.4%	-30.5%
Median Action[2]	503	1.8%	1.5%	**1.5%**	1.2%	1.3%	-1.0%
Pot Games	4244	29.9%	-2.7%	-1.1%	-1.3%	-1.9%	**9.9%**
Traveller's Dilemma	541	26.2%	**10.3%**	9.8%	9.4%	3.4%	2.7%
p-Beauty Contest[2]	519	585.4%	**49.9%**	40.8%	26.7%	-7.2%	-64.0%

Total Payoff and Percentage Improvement for Bionic Subjects using In-game Estimates[1]

	Observed Payoff	Ex-post %Improve	EWA Lite %Improve	EWA %Improve	Belief-based %Improve	Reinforcement with PV %Improve	QRE %Improve
Mixed Strategies	334	100.0%	13.0%	**15.9%**	14.1%	13.0%	1.0%
Patent Race	467	44.2%	**3.0%**	1.9%	2.9%	1.5%	1.2%
Continental Divide[2]	837	6.6%	**5.3%**	5.3%	5.2%	3.9%	-31.6%
Median Action[2]	503	1.8%	1.5%	**1.5%**	**1.5%**	1.2%	-1.0%
Pot Games	4244	29.9%	7.7%	**11.1%**	6.1%	9.3%	9.9%
Traveller's Dilemma	541	26.2%	11.0%	**11.1%**	9.5%	8.2%	7.6%
p-Beauty Contest[2]	519	585.4%	52.1%	**55.1%**	54.1%	-60.8%	-64.0%

Note 1: We assume that each bionic subject use the respective model to predict other's behavior and best responds with the strategy that yields the highest expected payoff.

Note 2: The expected value of each strategy in these games is computed with 1000 simulated instances for a given round due to high computational burden for actual derivation.

Table 6: Bionic Subjects with Better or Equal Payoffs using In-game and Out-of-game Estimates

Number of Bionic Subjects with Better or Equal Payoff using Out-of-game Estimates

	Total Subjects	EWA Lite % Improve	EWA % Improve	Belief-based % Improve	Reinforcement with PV % Improve	QRE % Improve
Mixed Strategies	8000.0%	**61.3%**	*53.8%*	51.3%	*57.5%*	43.8%
Patent Race	7200.0%	*54.2%*	*54.2%*	*52.8%*	**56.9%**	*52.8%*
Continental Divide	7000.0%	*94.3%*	**97.1%**	85.7%	15.7%	0.0%
Median Action	5400.0%	*87.0%*	**94.4%**	*83.3%*	*85.2%*	53.7%
Pot Games	8400.0%	47.6%	48.8%	48.8%	56.0%	**84.5%**
Traveller's Dilemma	5200.0%	**94.2%**	*90.4%*	84.6%	76.9%	55.8%
p-Beauty Contest	19600.0%	**68.9%**	61.2%	59.2%	53.1%	19.4%

Number of Bionic Subjects with Better or Equal Payoff using In-game Estimates

	Total Subjects	EWA Lite % Improve	EWA % Improve	Belief-based % Improve	Reinforcement with PV % Improve	QRE % Improve
Mixed Strategies	8000.0%	*70.0%*	*73.8%*	**75.0%**	*71.3%*	47.5%
Patent Race	7200.0%	*59.7%*	*54.2%*	**62.5%**	*52.8%*	*52.8%*
Continental Divide	7000.0%	**100.0%**	*98.6%*	*95.7%*	77.1%	0.0%
Median Action	5400.0%	*90.7%*	**96.3%**	**96.3%**	85.2%	53.7%
Pot Games	8400.0%	77.4%	**84.5%**	71.4%	*79.8%*	**84.5%**
Traveller's Dilemma	5200.0%	**96.2%**	**96.2%**	*90.4%*	*88.5%*	67.3%
p-Beauty Contest	19600.0%	**70.4%**	*68.9%*	*68.4%*	23.0%	19.4%

Table A.1: Parameter Estimates of Learning Models

EWA Lite [1]

	ϕ		κ		δ		N0 [3]		λ [5]	
Mixed Strategies	0.89	-	0.52	-	0.28	-	1.00	-	3.78	0.17
Patent Race	0.89	-	0.72	-	0.32	-	1.00	-	7.87	0.17
Continental Divide	0.69	-	0.77	-	0.69	-	1.00	-	4.46	0.18
Median Action	0.85	-	0.78	-	0.85	-	1.00	-	5.00	0.33
Pot Games	0.80	-	0.44	-	0.44	-	1.00	-	0.33	0.03
Traveller's Dilemma	0.63	-	0.84	-	0.63	-	1.00	-	4.99	0.20
p-Beauty Contest	0.58	-	0.82	-	0.58	-	1.00	-	2.11	0.05
Pooled	0.76	-	0.64	-	0.48	-	1.00	-	3.26	0.05

EWA

	ϕ		κ		δ		N0 [4]		λ	
Mixed Strategies	0.98	0.00	1.00	0.04	0.27	0.07	0.82	0.00	1.15	0.10
Patent Race	0.92	0.01	0.05	0.02	0.36	0.25	1.37	0.02	42.21	4.57
Continental Divide	0.74	0.03	1.00	0.02	0.73	0.09	0.25	0.00	3.98	0.30
Median Action	0.71	0.07	1.00	0.02	0.89	0.00	0.00	0.00	8.90	1.12
Pot Games	0.81	0.04	1.00	0.09	0.42	0.00	0.00	0.00	0.19	0.03
Traveller's Dilemma	0.77	0.02	1.00	0.02	0.53	0.07	0.62	0.00	3.53	0.22
p-Beauty Contest	0.36	0.02	0.00	0.04	0.78	0.05	1.56	0.00	3.44	0.00
Pooled [2]	0.78	0.01	0.99	0.01	0.49	0.00	0.01	0.00	2.95	0.02

Belief-based

	ϕ		κ [3]		δ [3]		N0 [4]		λ	
Mixed Strategies	1.00	0.00	0.00	-	1.00	-	72.59	2.18	43.34	0.24
Patent Race	1.00	0.00	0.00	-	1.00	-	27.78	141.08	85.09	0.01
Continental Divide	0.99	0.05	0.00	-	1.00	-	1.11	0.15	14.76	1.26
Median Action	1.00	0.00	0.00	-	1.00	-	5.48	8.58	75.84	0.08
Pot Games	0.98	0.04	0.00	-	1.00	-	0.49	0.17	0.91	0.13
Traveller's Dilemma	0.85	0.01	0.00	-	1.00	-	6.74	1.51	13.97	0.85
p-Beauty Contest	0.40	0.02	0.00	-	1.00	-	0.79	0.40	2.95	0.08
Pooled	0.81	0.02	0.00	-	1.00	-	5.36	0.49	11.59	0.48

Reinforcement with PV

	ϕ [3]		κ [3]		δ [3]		N0		λ	
Mixed Strategies	1.00	-	0.00	-	0.00	-	31.47	0.91	4.39	0.52
Patent Race	1.00	-	0.00	-	0.00	-	6.58	0.03	1.48	0.05
Continental Divide	1.00	-	0.00	-	0.00	-	1.81	0.13	2.59	0.03
Median Action	1.00	-	0.00	-	0.00	-	3.95	0.24	1.27	0.19
Pot Games	1.00	-	0.00	-	0.00	-	0.49	0.03	0.39	0.03
Traveller's Dilemma	1.00	-	0.00	-	0.00	-	3.92	0.03	3.19	0.03
p-Beauty Contest	1.00	-	0.00	-	0.00	-	1.78	0.01	0.22	0.01
Pooled	1.00	-	0.00	-	0.00	-	188.10	0.24	24.49	0.10

QRE

	ϕ [3]		κ [3]		δ [3]		N0 [3]		λ	
Mixed Strategies	0.00	-	0.00	-	0.00	-	0.00	-	10.27	0.08
Patent Race	0.00	-	0.00	-	0.00	-	0.00	-	4.60	0.04
Continental Divide	0.00	-	0.00	-	0.00	-	0.00	-	0.63	0.01
Median Action	0.00	-	0.00	-	0.00	-	0.00	-	20.00	0.00
Pot Games	0.00	-	0.00	-	0.00	-	0.00	-	0.04	0.00
Traveller's Dilemma	0.00	-	0.00	-	0.00	-	0.00	-	20.00	0.00
p-Beauty Contest	0.00	-	0.00	-	0.00	-	0.00	-	0.00	0.00
Pooled	0.00	-	0.00	-	0.00	-	0.00	-	7.78	0.01

Note 1: Average parameters across subjects and time.

Note 2: For all models except EWA Lite, a common set of estimates, except lambdas, is estimated for all games pooled.

Note 3: Fixed parameters

Note 4: N0 bounded by 1/(1-phi(1-kappa)).

Note 5: Payoffs in all games have been rescaled to USD equivalent. Average of game specific lambda is reported for pooled games.

Table A.2: Variations of EWA Lite Functional Values

Interquartile Range Across Time[1]

	ϕ		κ		δ	
	Median[3]	Range[4]	Median	Range	Median	Range
Mixed Strategies	0.91	0.0396	0.50	0.1027	0.30	0.1137
Patent Race	0.90	0.0200	0.73	0.0151	0.31	0.0078
Continental Divide	0.69	0.0806	0.78	0.0165	0.69	0.0806
Median Action	0.91	0.1875	0.81	0.1513	0.91	0.1875
Pot Games	0.81	0.0635	0.44	0.1329	0.44	0.0464
Traveller's Dilemma	0.62	0.0722	0.89	0.0926	0.62	0.0722
p-Beauty Contest	0.58	0.0375	0.88	0.0673	0.58	0.0375
Pooled	0.87	0.1218	0.57	0.2610	0.35	0.1512

Interquartile Range Across Subjects[2]

	ϕ		κ		δ	
	Median	Range	Median	Range	Median	Range
Mixed Strategies	0.89	0.0357	0.52	0.1504	0.27	0.1026
Patent Race	0.89	0.0133	0.74	0.2036	0.32	0.0044
Continental Divide	0.68	0.0820	0.78	0.0835	0.68	0.0820
Median Action	0.85	0.0000	0.79	0.0432	0.85	0.0000
Pot Games	0.78	0.0167	0.38	0.3247	0.43	0.0074
Traveller's Dilemma	0.62	0.0830	0.84	0.0121	0.62	0.0830
p-Beauty Contest	0.58	0.0156	0.82	0.0139	0.58	0.0156
Pooled	0.76	0.2694	0.82	0.2057	0.57	0.2664

Note 1: Average value for each time period is calculated, then the 25% and 75% percentile are used to derive the range.
Note 2: Average value for each subject is calculated, then the 25% and 75% percentile are used to derive the range.
Note 3: Overall medians on the average values calculated with respect to Note 1 and 2.
Note 4: The 25% and 75% percentile are respectively 0.5*Range below and above the median.

Figure 1: EWA's Model Parametric Space

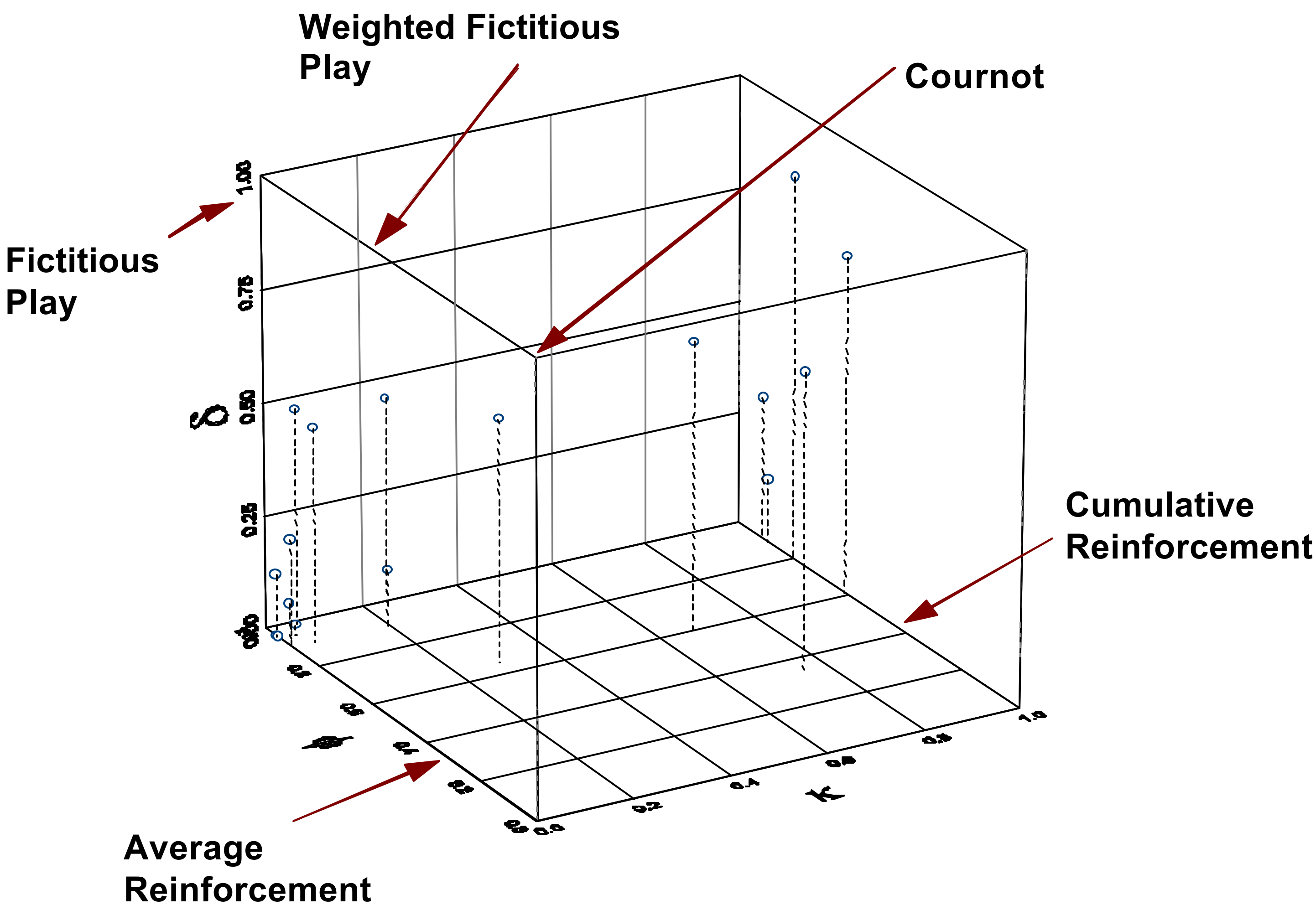

Figure 2: Parametric Plot of EWA Lite vs EWA

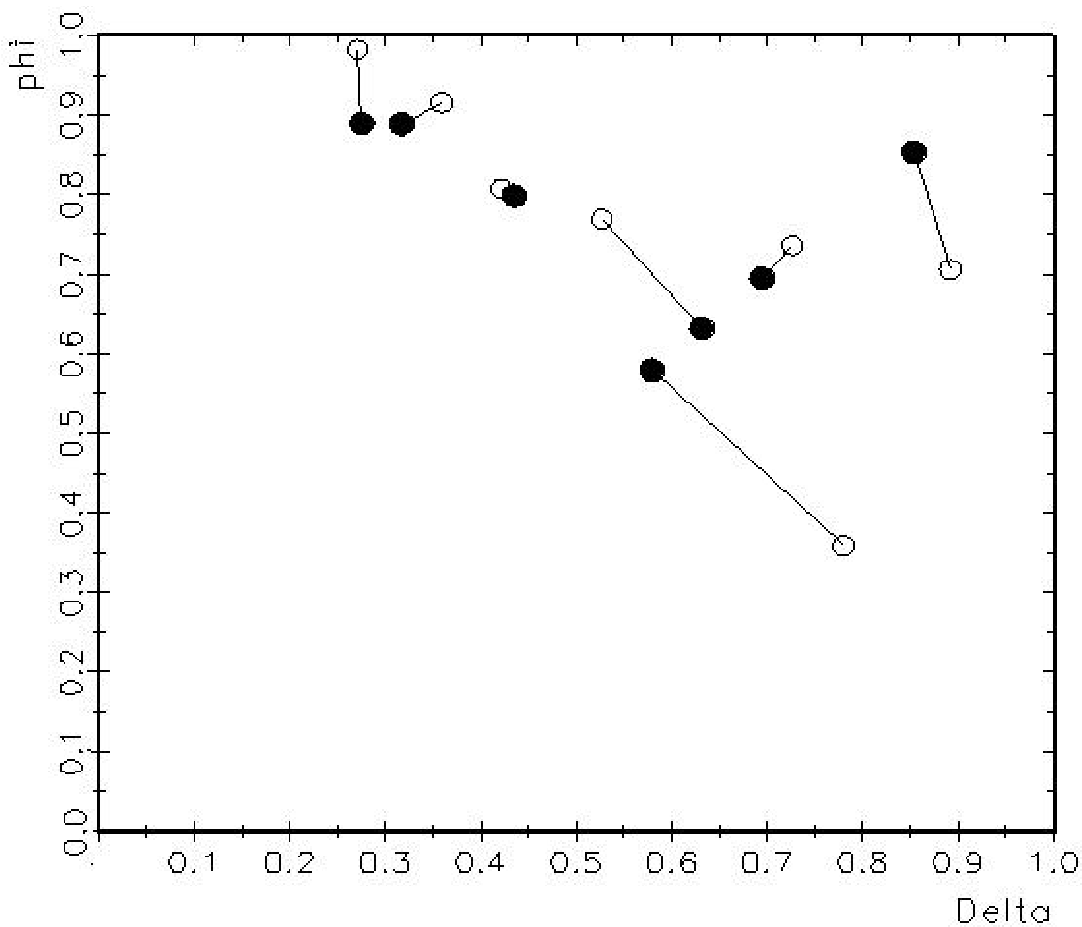

Figure 3 Transition Matrices for Patent Race

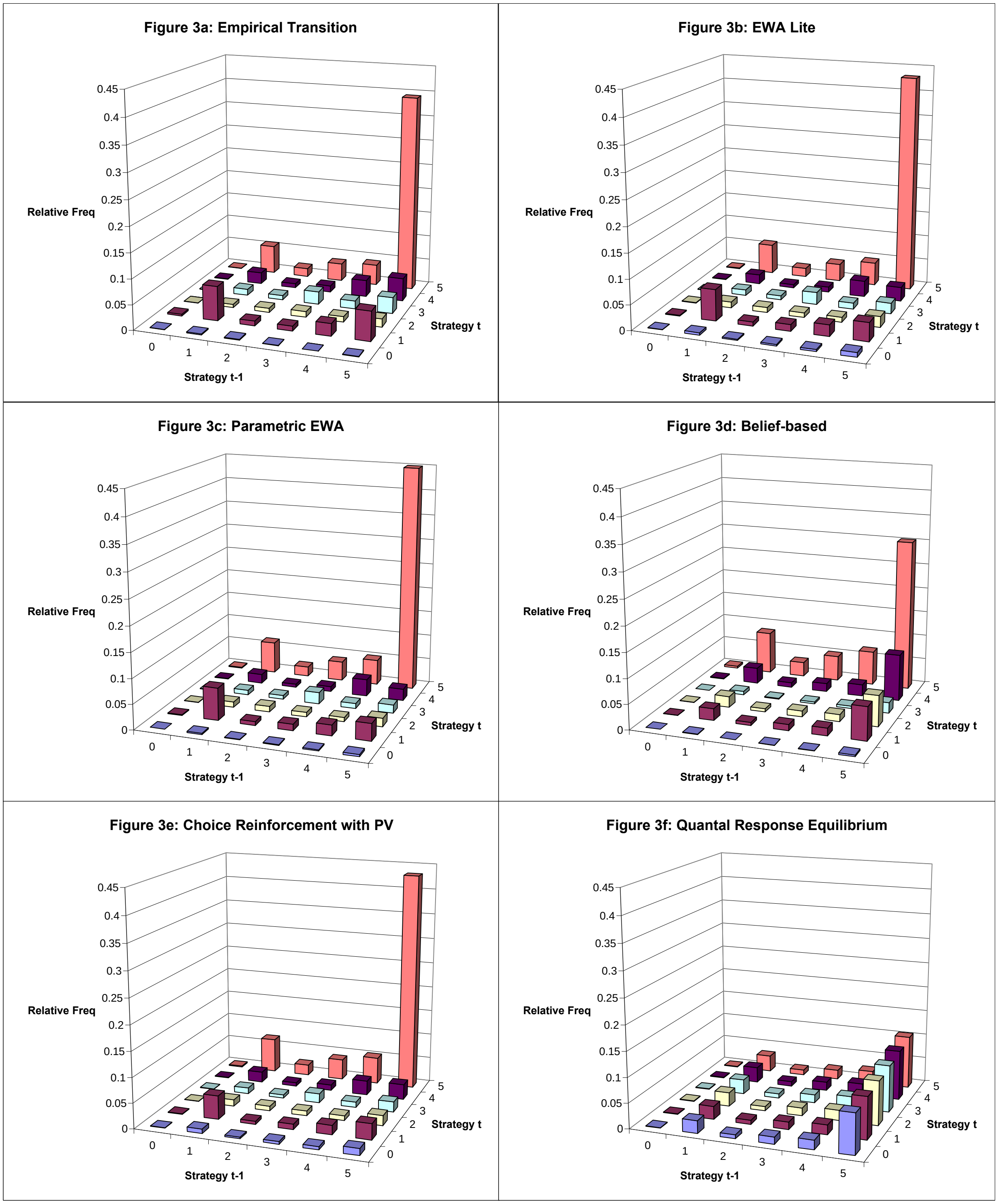

Figure 4: Empirical Frequency and Model Predictions for Continental Divide

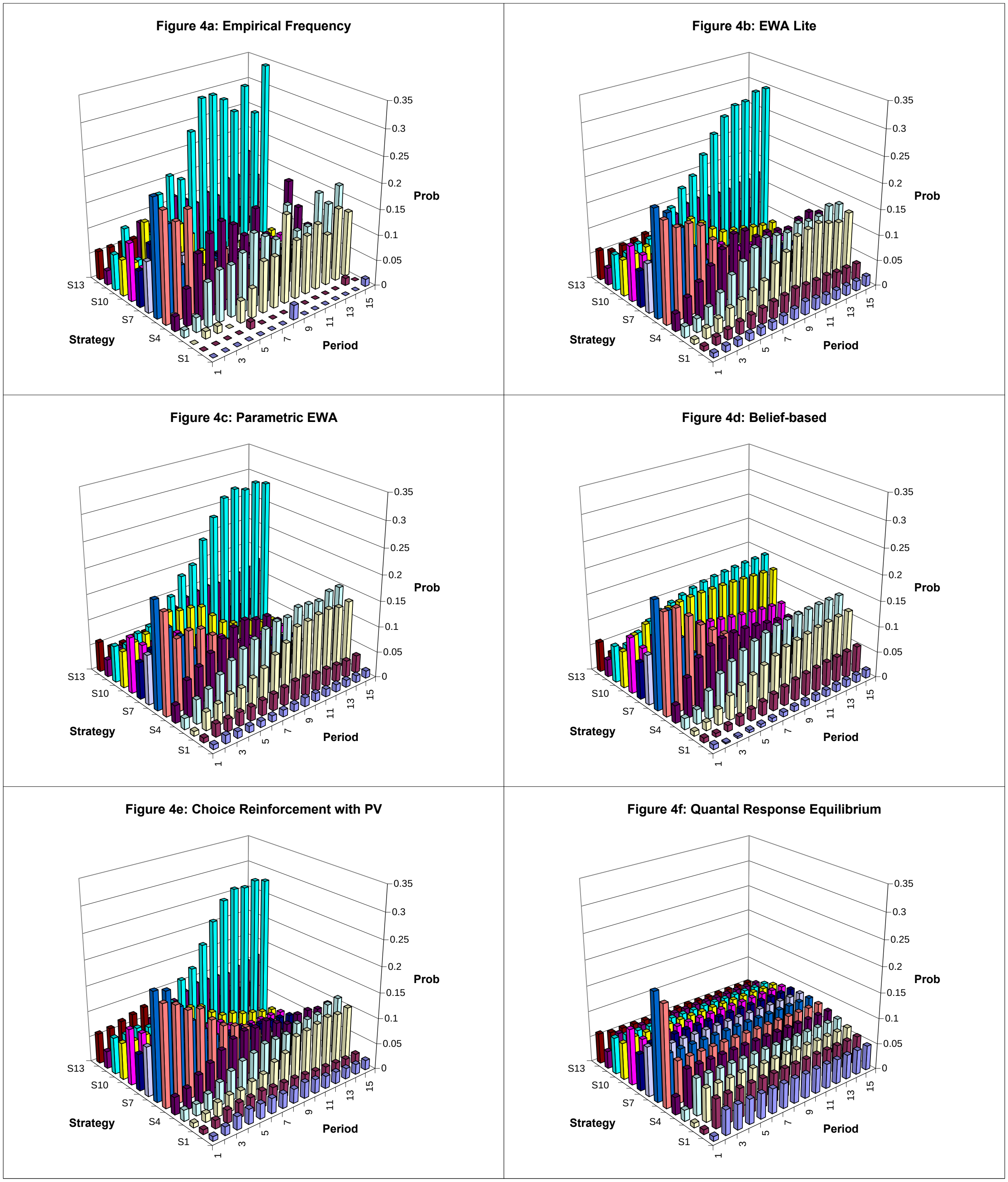

Figure 5: Empirical Frequency and Model Predictions for Traveller's Dilemma